



Universidad
Carlos III de Madrid

ESCUELA POLITÉCNICA SUPERIOR

Departamento de Teoría de la Señal y Comunicaciones



LOCALIZACIÓN Y SEGUIMIENTO DE PERSONAS EN ENTORNOS VIDEOVIGILADOS

INGENIERÍA DE TELECOMUNICACIÓN

Autor: Luis Martín - Cobos Blanco

Leganés, Enero 2015

Tutor: Francisco Javier González Serrano

Título: Localización y seguimiento de personas en entornos videovigilados

Autor: Luis Martin - Cobos Blanco

Director: Francisco Javier González Serrano

EL TRIBUNAL

Presidente: _____

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día ____ de _____
de 20__ en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de
Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE

AGRADECIMIENTOS

Doy gracias a mi familia por darme la oportunidad de llegar a este momento y conseguir un logro que supone un antes y después en mi vida.

Doy gracias a mis amigos, tanto del colegio como de la universidad, que me han dado muchos buenos momentos día tras día, complementando el duro trabajo previo que ha desembocado en la realización de este proyecto. Gracias especialmente a Alberto, que forma gran parte de este éxito al estar, a lo largo de estos 5 años, siempre atento a todo lo que también me afectaba y acompañarme en lo que ha sido probablemente la mejor decisión de mi vida hasta el momento, estudiar en Suecia.

Por último, gracias a Fran, profesor que me ha enseñado y formado durante todos los 5 años de carrera, que formó parte de nuestra aventura en Suecia y que ha sido un tutor que me ha guiado y asesorado perfectamente durante el periodo que ha durado este proyecto que ha sido todo un éxito.

RESUMEN

El cálculo de trayectorias ha sido y es una aplicación muy usada en diferentes campos. Desde saber cuánto ha corrido un jugador de fútbol hasta hacer seguimiento de personas a lo largo de un centro comercial.

Es un campo en continua evolución y de gran interés puesto que hace uso de técnicas como la detección de movimiento, que son usadas en multitud de aplicaciones como pueden ser la seguridad (aparcamientos), entretenimiento (kinect en la Xbox), iluminación (encender luces cuando se encuentra en la habitación), frenado automático de coches al detectar personas delante de ellos, etc.

Con este proyecto se pretende conseguir calcular las trayectorias seguidas por las personas dentro del campo de visión de una cámara. Como la cámara nos da información en dos dimensiones y el espacio real son tres dimensiones, necesitaremos la ayuda de un elemento extra que nos proporcione la información que nos falta.

Se realizará un cálculo de la trayectoria seguida por la persona gracias a una cámara, un sensor de distancia incluido en la persona y un ordenador para realizar el procesado.

Sabemos que debería ser posible obtener la posición 3D de la persona al obtener un sistema de tres ecuaciones con tres incógnitas, dos proporcionadas por la cámara y una por un sensor de distancia.

Este documento consistirá en una descripción del estado del arte, el diseño concreto llevado a cabo, unos estudios de los resultados y una conclusión final.

Palabras clave:

Detección de movimiento, calibración, persona, trayectoria, sensor, cámara, iluminancia, croma, detección de bordes, detección de personas, erosión, distorsión, filtro de Kalman, filtro de mediana, Componentes conectados.

ABSTRACT

Tracking people has been and is widely used in different application fields. How long or the distance a football player has run, every position a person has been inside the store or tracking a person over a mall could be different applications.

It is an evolving field with a lot of interest due to it makes use of techniques like motion detection, which is used in many applications such as security, entertainment, lighting, comfort, automatic braking car to detect people in front of them, etc.

The goal of this project is to be able to track people in surveillance environments using the minimum number of elements, in this case, a camera.

In order to achieve this aim, a calculation of the person path using a camera, a wearable or portable distance sensor and a computer is performed.

We know it should be possible, at first glance, to obtain a system of three equations with three variables collected by different systems. Two of the equations will be retrieved by the camera and the remaining one will be by the distance sensor.

This paper consists of a description of the state of the art, detailed design and architecture of our system, evaluation of the results and a final conclusion.

Keywords

Motion detection, calibration, people classification, career, sensor, camera, luminance, chrominance, edge detection, detection of people, Gaussian distribution , erosion, distortion, Kalman filter, median filter, Connected components.

ÍNDICE DE CONTENIDOS

Capítulo 1: INTRODUCCIÓN.....	1
1.1 Motivación	1
1.2 Objetivos y fases.....	2
1.3 Aplicaciones de interés	6
Capítulo 2: ESTADO DEL ARTE	11
2.1 Diferentes aproximaciones	11
2.1.1 Dos o más cámaras.....	11
2.1.2 Una cámara y varios sensores láser	13
2.1.3 Aproximación propuesta: Una cámara y un sensor de distancia	14
2.2 Fases para solucionar el problema.....	15
2.2.1 Calibración.....	15
2.2.2 Segmentación del fondo	16
2.2.3 Clasificación de los objetos detectados	19
2.2.4 Fusión de sensores	20
Capítulo 3: DISEÑO	23
3.1 Perspectiva general.....	23
3.2 Calibración.....	23
3.2 Segmentación del fondo	27
3.2.1 Mixture of Gaussians background extraction	29
3.2.2 Luminance and chrominance distortion background extraction	32
3.2.3 Edge background extraction.....	36
3.2.4 Procesado posterior de la imagen.....	37
3.3 Clasificación de personas	42
3.4 Fusión de sensores	46
Capítulo 4: ESTUDIO Y EVALUACIÓN DE LOS RESULTADOS	54
4.1 Calibración.....	54
4.2 Segmentación del fondo	54
4.2.1 Mixture of Gaussians background extraction	54
4.2.2 Edge background extraction	55

4.2.3	Luminance and chrominance distortion background extraction	56
4.3	Clasificación de personas	60
4.4	Fusión de sensores	62
Capítulo 5: CONCLUSIÓN.....		65
REFERENCIAS.....		70
ANEXO		74

Índice de figuras

Figura 1. Escenario de aplicación

Figura 2. Interpretación de proyecciones

Figura 3. Representaciones perro detectado

Figura 4. Interpretación de proyecciones con oclusión

Figura 5. Aplicaciones de interés

Figura 6. Aplicación en aparcamientos

Figura 7. Detección y seguimiento de multitudes en un centro comercial

Figura 8. Detección y caracterización de jugadores

Figura 9. Captura de movimiento 1

Figura 10. Captura de movimiento 2

Figura 11. Transformación sobre el plano común de dos cámaras

Figura 12. Plano de tierra común después del uso de homografías 2D

Figura 13. Arquitectura del sistema de detección basada en láseres y cámaras

Figura 14. Esquema típico de un sistema de extracción de fondo

Figura 15. Oclusión

Figura 16. Esquema de un sistema de seguimiento

Figura 17. Etapas principales del proyecto

Figura 18. Diagrama de flujo de la sección de calibración

Figura 19. Representación del vector de traslación y matriz de rotación

Figura 20. Imagen calibración 1

Figura 21. Imagen calibración 2

Figura 22. Imagen calibración 3

Figura 23. Imagen calibración 4

Figura 24. Imagen calibración 5

Figura 25. Imagen calibración 6

Figura 26. Diagrama de flujo de la extracción del fondo

Figura 27. Peso de las gaussianas sobre un pixel

Figura 28. Separación de la luz en sus componentes de luminancia y croma

Figura 29. Variación de la distorsión del brillo

Figura 30. Variación de la distorsión de la croma

Figura 31. Imagen de referencia para los bordes verticales

Figura 32. Imagen de referencia para los bordes horizontales

Figura 33. Grafo de representa tres caminos conectados

Figura 34. Tipos de filtros de mediana y su resultado

Figura 35. Uso del filtro de mediana para corregir una imagen corrompida por la mala recepción

Figura 36. Máscara

Figura 37. Extracción de la frontera mediante el uso de la erosión

Figura 38. Elemento estructural usado en la dilatación

Figura 39. Uso de la dilatación para rellenado de figuras

Figura 40. Imagen capturada

Figura 41. Obtención del movimiento con Connected Components

Figura 42. Obtención del movimiento con todo el post procesado.

Figura 43. Diagrama de flujo de la clasificación de personas

Figura 44. Imagen de referencia

Figura 45. Obtención del centro de la figura objetivo

Figura 46. Obtención del eje de mayor varianza y su perpendicular (PCA)

Figura 47. Proyección horizontal

Figura 48. Proyección vertical

Figura 49. Proyección vertical de referencia

Figura 50. Proyección horizontal de referencia

Figura 51. Diagrama de flujo del cálculo de la posición 3D

Figura 52. Detección realizada

Figura 53. Punto en 3D correspondiente a la detección de la figura 40

Figura 54. Trayectoria en 3D obtenida

Figura 55. Funcionamiento del filtro de Kalman

Figura 56. Esquema completo de la operación del Filtro de Kalman.

Figura 57. Uso del filtro de Kalman

Figura 58. Representación de la trayectoria ideal, trayectoria obtenida con y sin filtro de Kalman

Figura 59. Rendimiento del algoritmo

Figura 60. Rendimiento de este algoritmo

Figura 61. Rendimiento de algoritmo basado en Gaussian

Figura 62. Rendimiento del algoritmo basado en luminancia y croma

Figura 63. Imágenes referencia cálculo rendimiento

Figura 64. Plantillas referencia cálculo de rendimiento

Figura 65. Rendimiento total fase detección fondo 1

Figura 66. Rendimiento total fase detección fondo 2

Figura 67. Rendimiento total fase detección fondo 3

Figura 68. Rendimiento total fase detección fondo 4

Figura 69. Clasificación de una persona sin otros elementos detectados

Figura 70. Clasificación de una personas con más elementos detectados

Figura 71. Clasificación de dos personas con otros elementos detectados

Figura 72. Trayectoria obtenida y error en mínimos cuadrados

Figura 73. Estudio de las consecuencias del error introducido por el sensor de distancia

Índice de tablas

Tabla 1. Métodos de modelado del fondo

Tabla 2. Clasificación de los píxeles

Tabla 3. Cálculo de la confianza para un píxel

Tabla 4. Ecuaciones de actualización de tiempo del filtro de Kalman discreto.

Tabla 5. Ecuaciones de actualización de mediciones del filtro de Kalman discreto

Tabla 6. Ecuaciones de reasignación de variables

Tabla 7. Rendimiento de los algoritmos de segmentación del fondo

Capítulo 1: INTRODUCCIÓN

1.1 Motivación

La información, sea del tipo que sea, es una fuente de conocimiento que permite actuar de forma que se tenga un porcentaje mayor de éxito en el campo en el que se aplica. La información es una materia prima muy valiosa que es usada por todas las empresas del mundo. Saber dónde colocar los objetos que interesa especialmente vender en la zona en la que las personas se paran más tiempo o qué objeto colocar en qué escaparate son preguntas que este proyecto ayuda a responder.

Con este proyecto se pretende estudiar la capacidad de una infraestructura comúnmente ya desplegada en la mayoría de lugares de aplicación, es decir, donde hay videovigilancia, para calcular la trayectoria seguida por personas y obtener información muy valiosa de ella.

El punto a resaltar que caracteriza este proyecto es la capacidad de dar un gran valor añadido a un despliegue que ya se ha realizado. Además, hará de punto de partida común de diferentes aplicaciones que después se quieran desarrollar.

Algunos ejemplos de estas aplicaciones podrían ser: un procesado para determinar hacia donde está mirando el individuo en el caso de situar objetos en los escaparates, un procesado para saber cuánto tiempo pasa un individuo en cada fila de estanterías para situar los objetos según convenga, etc.

1.2 Objetivos y fases

El objetivo principal de este proyecto es conseguir obtener trayectorias de personas haciendo uso de la infraestructura de vigilancia ya desplegada en los lugares de aplicación. En este caso, se ha escogido una cámara como el elemento que formará la infraestructura al ser el elemento más básico que podría formarla.

Al tener una referencia del lugar en dos dimensiones que proporciona la cámara (la imagen de la pantalla) de un entorno en tres dimensiones (lugar observado), es necesario un elemento extra que nos aporte la información que falta. Este elemento extra será transportado por las personas por lo que es necesario una ligera colaboración por su parte.

En la siguiente figura se representa el escenario de referencia que se ha tenido en cuenta a la hora de realizar el estudio en este proyecto. Además, se puede apreciar la infraestructura anteriormente definida que debería ser suficiente para segmentar el fondo, detectar personas y conseguir saber, finalmente, la trayectoria seguida por cada persona.

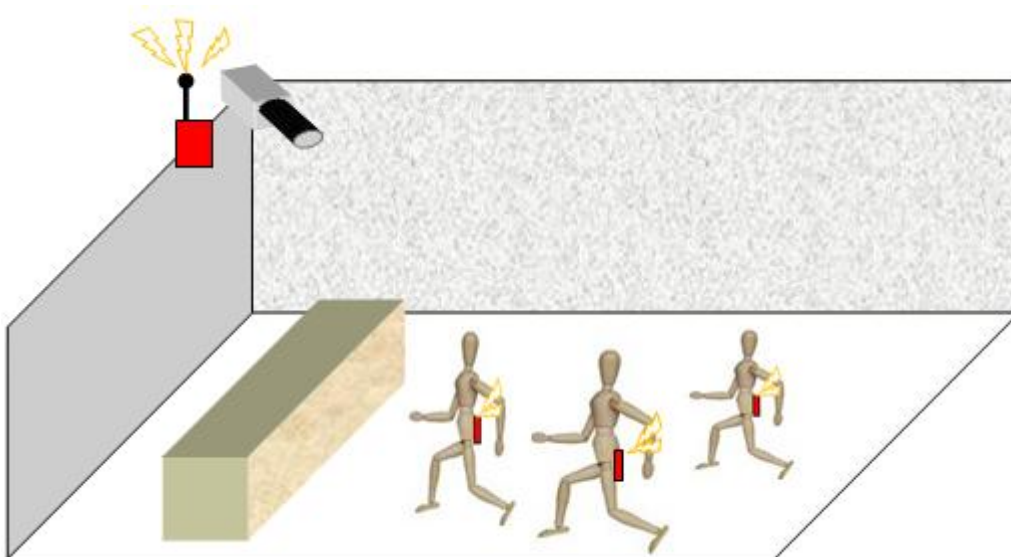


Figura 1. Escenario de aplicación

La segmentación del fondo y la correcta clasificación del objetivo en movimiento detectado es el principal foco de atención que ha requerido el proyecto. Esto es debido a que para poder llevar a cabo un seguimiento correcto y obtener la trayectoria, es necesario conseguir detectar a los objetivos lo más definidos posible durante toda su trayectoria a pesar de tener que lidiar con problemas como la oclusión con otras personas, con el mobiliario, etc.

Para extraer el fondo, se hace uso de la principal característica de un pixel (aplicándolo a toda la imagen), es decir, el color que tiene y su división en los componentes de luminancia y croma. También se comprobarán cambios en los bordes de los objetos que forman el fondo provocados por elementos moviéndose delante de ellos.

La detección de personas se realiza mediante comparación de las proyecciones de los puntos pertenecientes a la figura detectada sobre dos ejes con unas proyecciones de referencia [1]. Estas proyecciones de referencia, que se han obtenido previamente en una fase de entrenamiento, son calculadas haciendo uso de detecciones de objetivos se sabía que eran personas.

A continuación se va a explicar el porqué de considerar los puntos indicados anteriormente como los más críticos.

En la siguiente figura, las primeras dos proyecciones (vertical y horizontal) serían un ejemplo de proyecciones de referencia obtenidas en la fase de entrenamiento (para crear las de referencia) y que se corresponden, con ligeras variaciones, con las obtenidas en casos en los que la figura está detectada de forma bien definida.

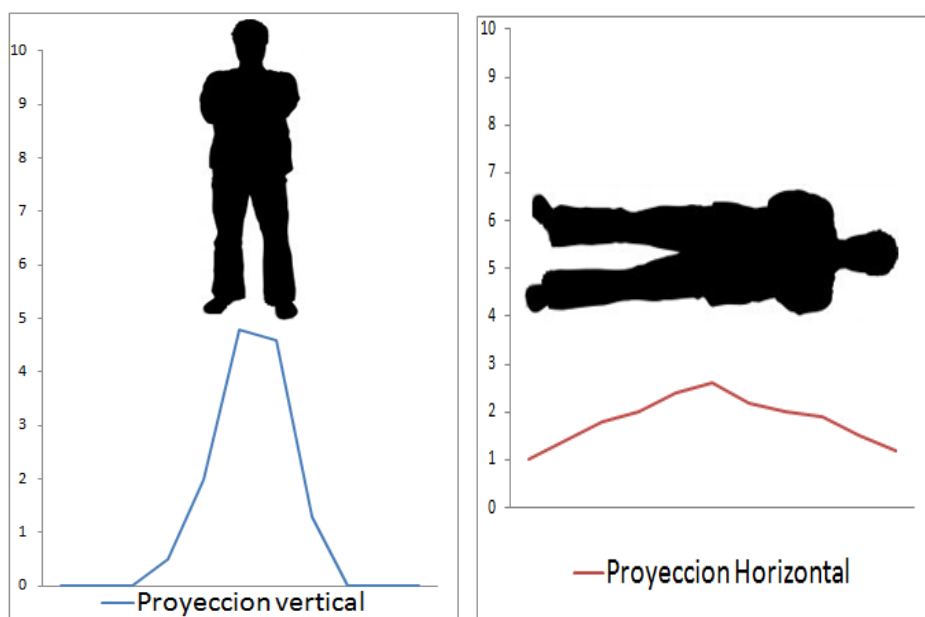


Figura 2. Interpretación de proyecciones

Como se puede comprobar en la siguiente figura, cuando se detecta un elemento en movimiento como por ejemplo, un perro, las proyecciones varían de forma que permiten identificar ese elemento como no persona.

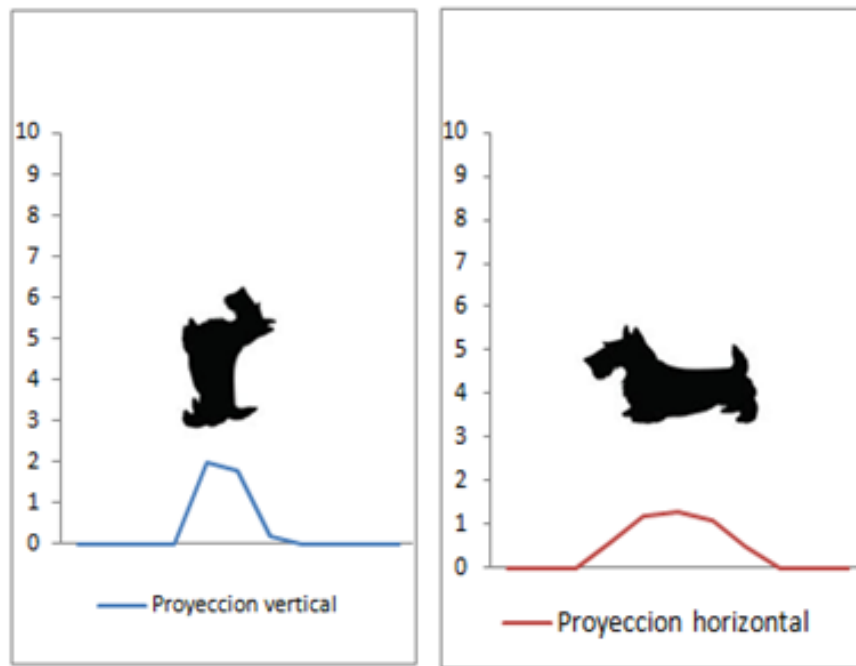


Figura 3. Representaciones perro detectado

En el caso de oclusión parcial, como la provocada por un banco de piedra, que se refleja en la siguiente figura, se puede observar como las proyecciones se ven deterioradas. En función del tipo de oclusión, puede variar la altura y la anchura de las proyecciones, solo la altura, solo la anchura, etc.

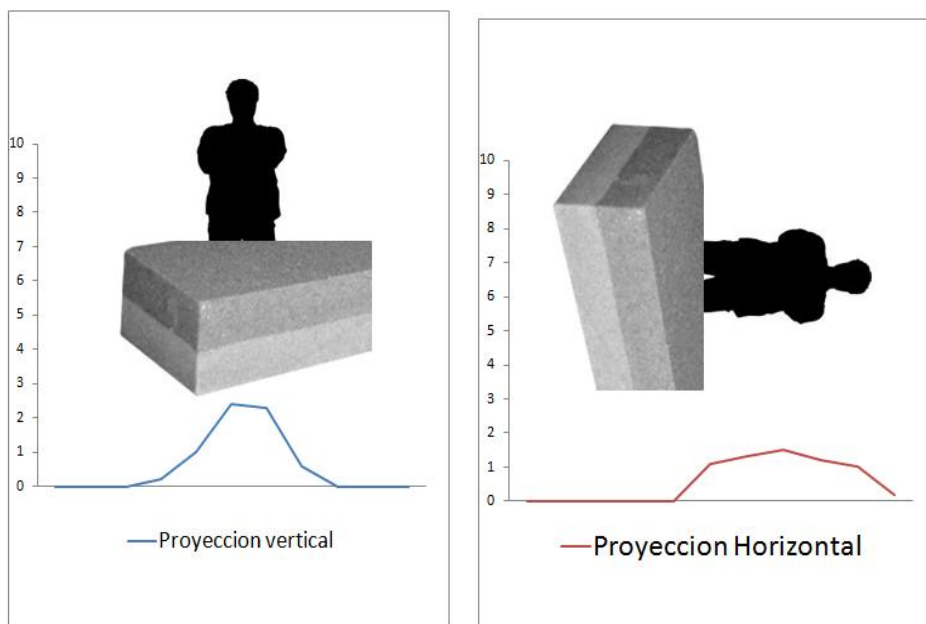


Figura 4. Interpretación de proyecciones con oclusión

Por lo tanto, conseguir una silueta del elemento en movimiento lo mejor definida posible y saber corregir los problemas que provocan la modificación de estas proyecciones es clave para conseguir llegar al objetivo final, puesto que un elemento

en movimiento que se detectara como no persona no sería de valor para este sistema y no se continuaría con su procesado.

El elemento extra que nos aportará la información que falta para resolver el problema puede ser cualquier dispositivo que sea capaz de darnos la distancia desde la persona a un receptor. Desde medir la potencia de la señal Wifi del móvil de las personas (convertir después esa potencia a distancia) hasta sensores de distancia que puedan estar incluidos en cualquier elemento que las personas puedan transportar son factibles de ser válidos como el elemento extra.

En nuestro caso, se ha escogido usar un sensor de distancia por ser un método poco invasivo y que puede ser usado en varios tipos de negocios diferentes al ser posible incluirlo en pulseras, carros de supermercado, lectores de códigos de barras para ver precios, etc.

A continuación, se hace una breve descripción de las fases que se van a encontrar a lo largo de la memoria de este proyecto.

Estudio del estado del arte:

En el segundo capítulo del presente proyecto, se hará una introducción a las posibles arquitecturas en las que se puede enfocar el problema para darle solución.

Además, se incluirá el conjunto de algoritmos que se pueden usar para solucionar cada uno de los pasos a seguir para resolver el problema planteado y que se podría aplicar en cualquiera de las arquitecturas anteriormente mencionadas.

Diseño e implementación de algoritmos y técnicas:

Una vez conocida la problemática, se realizará un diseño que permita integrar las técnicas elegidas y conformar un sistema que sea capaz de resolver los problemas a los que se enfrenta este Proyecto Fin de Carrera y cumplir el objetivo final, ser capaces de obtener la trayectoria seguida por personas.

Estudio y evaluación de los resultados

Mediante un conjunto de pruebas, se realizará una evaluación de los resultados que se obtienen con cada algoritmo y se tomará la decisión de cuál es el más idóneo.

También, se evaluará el rendimiento total del sistema presentado calculando el error de la trayectoria obtenida con respecto a la trayectoria ideal. Se analizará que

efecto tiene una recepción errónea de la distancia por parte del sensor sobre el cálculo de la trayectoria.

El análisis de los resultados nos permitirá extraer unas conclusiones finales y determinar si este trabajo es apropiado para su integración en sistemas comerciales.

1.3 Aplicaciones de interés

En este apartado se va a realizar una introducción a las posibles aplicaciones que pueden hacer uso de la arquitectura y algoritmia que se ha definido y usado en este proyecto.

<i>APLICACIONES</i>	
CAMPO GENERAL	ÁREA ESPECÍFICA
<ul style="list-style-type: none"> ✓ Sistemas de vigilancia inteligentes ✓ Análisis de datos ✓ Captura de movimiento 	<ul style="list-style-type: none"> • Control de accesos • Estacionamientos • Supermercados y tiendas • Tráfico • Creación de estadísticas • Personalización de entrenamientos • Animación basada en movimiento

Figura 5. Aplicaciones de interés

✓ Sistemas de vigilancia inteligentes

En este contexto, se entiende por inteligente a un sistema que hace más que detección de movimiento, ya que es un proceso que es propenso a dar falsas alarmas con el movimiento de objetos que no serían de interés como podrían ser animales, hojas, etc.

La primera necesidad que debería de cubrirse es saber identificar si ese objeto en movimiento es una persona o no. A continuación, se podría realizar un reconocimiento de caras como propósito de seguridad (acceso remoto) o de

seguimiento a través de varias cámaras (la cara reconocida estaría en una base de datos común a la estructura de cámaras), se podría identificar la pose de la persona detectada o cuánto tiempo lleva en la misma posición como indicadores que nos den cierto nivel de sospecha y la posibilidad de hacer saltar una notificación que muestre ese comportamiento sospechoso.

Saber la pose de una persona sería de gran ayuda para los sistemas de vigilancia de un aparcamiento ya que permitiría saber si una persona está buscando algo en los coches, es decir, con una postura inclinada y encogida.

En la siguiente imagen se representa cómo quedaría la aplicación de este proyecto a una situación tan común como un aparcamiento, donde se detecta a una persona entre todos los coches. Después, sería necesario obtener su postura para hacer saltar la notificación o no.



Figura 6. Aplicación en aparcamientos (imagen obtenida de <http://www.perimeterprotectivesystems.com/content/parking-lots>)

En la siguiente imagen se puede apreciar cómo se es capaz de obtener la trayectoria seguida por cada persona detectada en la imagen que permitiría saber cuáles son los locales más visitados y así, por ejemplo, establecer los precios de la publicidad en función del lugar del centro comercial donde se quiera colocar.



Figura 7. Detección y seguimiento de multitudes en un centro comercial
(imagen obtenida de
<http://www.perimeterprotectivesystems.com/content/parking-lots>)

✓ Análisis de datos

Al conseguir la trayectoria de una persona y todas las posiciones por las que ha pasado, es posible obtener información útil para la creación de estadísticas y generación de informes y complementar, por ejemplo, la retransmisión de un partido de fútbol. Es bastante común saber, por ejemplo, cuánto ha recorrido un jugador de fútbol para después poder mostrar estas estadísticas a los telespectadores.

Combinando el efecto de dos cámaras o más, sería posible conseguir saber la trayectoria y distancia recorrida por un jugador de fútbol o baloncesto aplicando la misma filosofía seguida en este proyecto. Hay que tener en cuenta que la capacidad de mantener identificado a cada jugador en este caso es muy sencilla al realizarse a través del número que llevan en la camiseta.

En la siguiente imagen se ha ilustrado esta posible aplicación haciendo uso de la filosofía seguida en este proyecto. En ella se puede comprobar cómo cada jugador se encuentra identificado por un color y se hace un seguimiento de por qué zona del campo se ha movido y por lo tanto, la distancia recorrida.

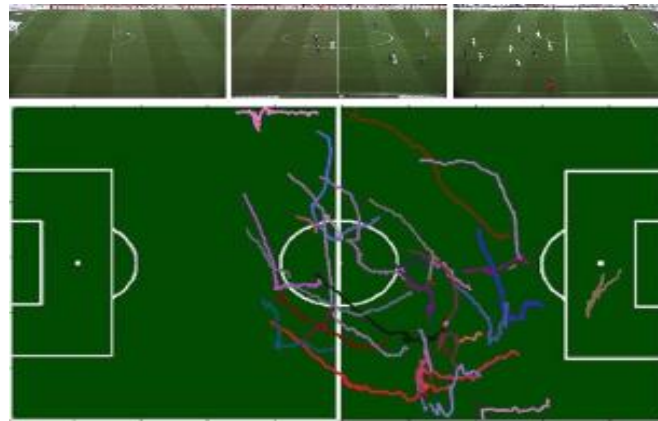


Figura 8. Detección y caracterización de jugadores (imagen obtenida de <http://cis.eecs.qmul.ac.uk/MScCV.html>)

✓ Captura de movimiento

Para poder crear animaciones para el rodaje de películas o diseño de videojuegos, sería posible hacer uso de la arquitectura propuesta en este proyecto cambiando el elemento extra (el sensor de movimiento) por varios sensores distribuidos a lo largo del cuerpo. Fusionando toda la información recogida por la cámara y los sensores, se podría ser capaz de seguir el movimiento de las partes del cuerpo que ha realizado la persona.

En las siguientes imágenes se quiere representar cómo esta metodología es usada actualmente a la hora de realizar una buena animación y tiene gran importancia en el campo de la ficción.

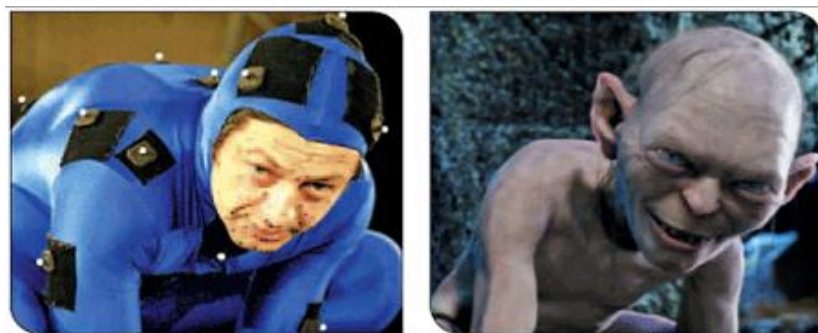


Figura 9. Captura de movimiento 1 (imagen obtenida de <http://ccorner.duke4.net/the-idea-of-capturing-motion/>)



Figura 10. Captura de movimiento 2
(<http://spectrum.ieee.org/computing/software/weta-digital-reverse-engineers-the-human-face>)

Capítulo 2: ESTADO DEL ARTE

En este segundo capítulo se va explicar qué arquitecturas se han implementado que puedan resolver el problema planteado. Además, se va a introducir las posibles soluciones a cada tarea que es necesario que realice cualquier arquitectura elegida, es decir, calibración, segmentación de fondo, clasificación de personas y fusión de sensores.

2.1 Diferentes aproximaciones

2.1.1 Dos o más cámaras

Esta aproximación crea una red de varias cámaras situadas a lo largo del escenario de forma que se cubra el mayor campo de visión posible y se intente evitar el mayor número posible de zonas que puedan provocar oclusión [2].

Las principales dos ventajas que presenta esta arquitectura son:

1. Es capaz de resolver los problemas relacionados con la oclusión mucho mejor que cualquier otra aproximación gracias a que se tiene mucha más visión de la escena y una oclusión sobre una cámara no sería tan determinante en la detección final.
2. Los elementos que forman esta arquitectura son pasivos, es decir, no necesita colaboración de los individuos para poder obtener la trayectoria.

Además, gracias al incremento del número de cámaras, es capaz de cubrir lugares de grandes dimensiones y es una solución escalable. Añadir más cámaras simplemente supondría usar los mismos algoritmos adaptándolo para una cámara mas, que en general, es bastante simple.

El principal problema a su vez que se obtiene es que cada cámara tiene una perspectiva y percepción de las distancias diferente y por tanto es necesario, mediante algún tipo de transformación o procesado, conseguir que todas las cámaras al final, se refieran, en términos de la posición del objetivo y distancia recorrida, sobre un espacio 3D común. También, esta arquitectura presenta el problema que todo el procesado de mantener identificada cada persona del resto de personas (para calcular la trayectoria) es necesario que lo hagan las cámaras puesto que no hay otro elemento posible que

permita identificar de forma inequívoca cada persona detectada. Esta tarea es complicada y requiere una capacidad de procesamiento importante.

La solución más típica que podemos encontrar es la siguiente:

- Seguimiento de objetos basado en proyecciones con homografías

En este algoritmo, se usan homografías para transformar cada posición del espacio 3D del que tiene perspectiva una cámara en particular en el espacio 3D común que tendrán todas las cámaras y que será el de referencia [2].

En términos matriciales, la transformación realizada se puede describir de la siguiente manera:

$$X = H_v x \quad (1)$$

donde H_v es una matriz de tamaño 3x3 homogénea. Las coordenadas de ambos espacios 3D se representan por los vectores $X = (X, Y, W)^T$ y $x = (x, y, 1)^T$

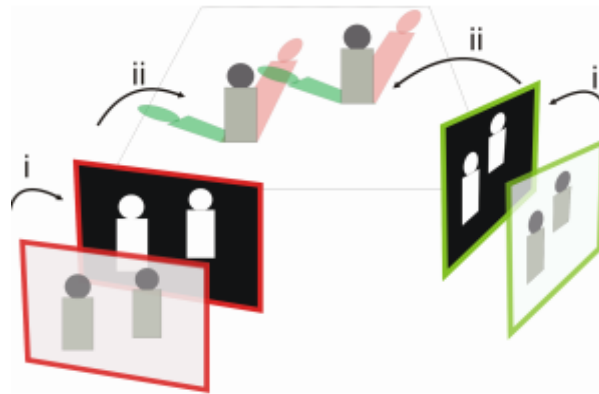


Figura 11. Transformación sobre el plano común de dos cámaras (imagen obtenida de [2])

En la figura anterior se puede observar cómo diferentes cámaras provocan diferentes perspectivas de la escena. Se realiza detección de movimiento (paso i) y después, mediante homografías, se pasa del plano 3D de la cámara al plano 3D común para todas las cámaras (paso ii)

Una vez se tienen todos los puntos transformados, es necesario evaluar la posición más probable en la que se encuentra el objetivo.

En la siguiente figura se puede observar cómo mediante este método y varias cámaras, se obtiene un conjunto de puntos donde es muy probable que se encuentren los objetivos.

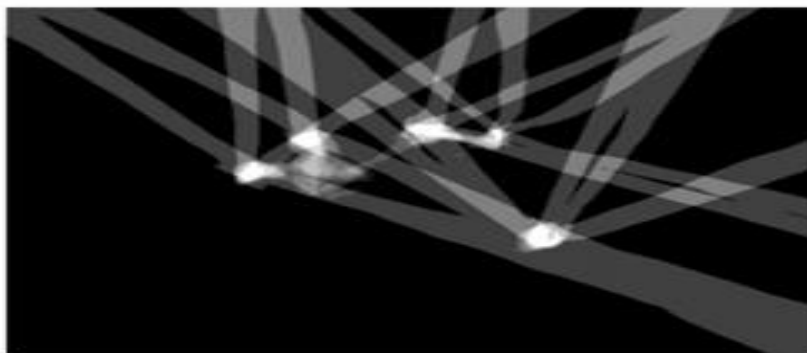


Figura 12. Plano de tierra común después del uso de homografías 2D (imagen obtenida de <http://humanmotion.rutgers.edu/HM/W02.008.pdf>)

2.1.2 Una cámara y varios sensores láser

Al utilizarse esta aproximación, se crea una red de sensores que incluye una cámara y varios sensores láser distribuidos a lo largo del lugar de aplicación intentando cubrir, teniendo en cuenta el alcance de cada láser, la mayor área posible. Los métodos de detección y seguimiento basados en detectores láser pueden dar una detección bastante fiable de un humano en varios escenarios. Además, esta aproximación también incluye la ventaja de que sus sensores son pasivos y no necesitan colaboración de los objetivos para llevar a cabo su función. Son insensibles a la variación de iluminación y el procesamiento de datos de los dispositivos láser no consume mucho tiempo de procesamiento. Un solo escáner láser es capaz de hacer seguimiento de varias personas [3].

Cada sensor de esta aproximación está controlado por un ordenador, llamado ordenador cliente. Todos los ordenadores clientes (láseres y cámara) están conectados mediante una red a un ordenador servidor que sincroniza los datos obtenidos de los sensores y contiene los datos de todos los ordenadores clientes.

Para sincronizar los datos, los relojes de los equipos cliente locales son corregidos periódicamente con el reloj del equipo servidor. Cada vez que un sensor toma datos se registra el tiempo en el que se ha hecho y después, el ordenador servidor es el que se encarga de juntar toda la información basándose en ese registro del tiempo en el que se han obtenido los datos.

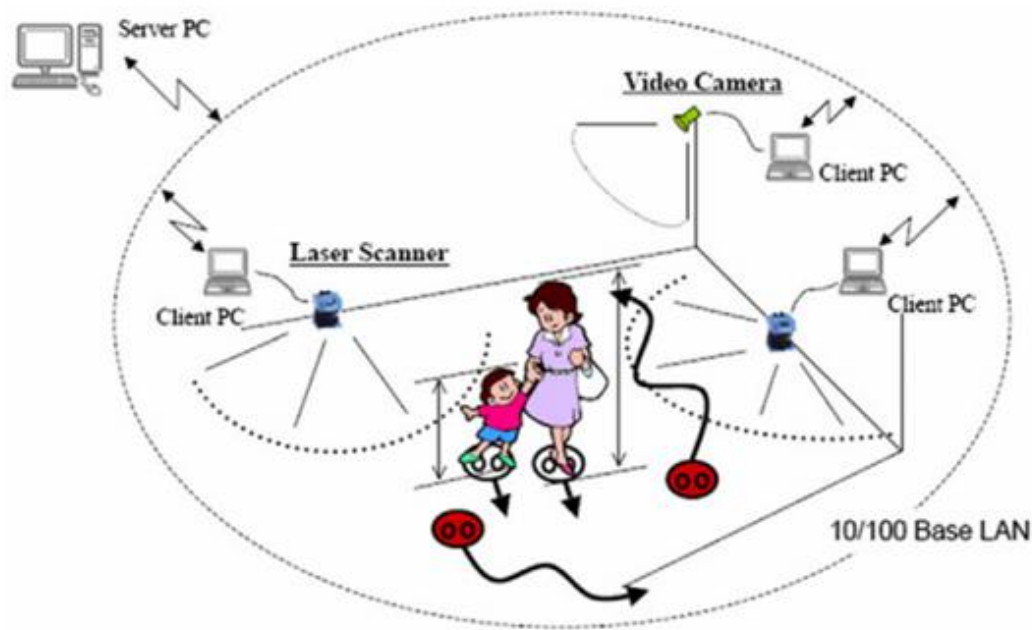


Figura 13. Arquitectura del sistema de detección basada en láseres y cámaras (imagen obtenida de [3])

En la figura anterior se puede ver un ejemplo de arquitectura usada por este tipo de métodos y las diferentes interconexiones que se pueden tener para obtener el máximo rendimiento.

El principal problema que tiene esta arquitectura es que los láseres no son capaces de saber identificar a una persona del resto de objetivos a lo largo de su campo de acción. Esto representa un gran problema a la hora de obtener la trayectoria seguida por cada persona cuando se cruzan ya que todo este procesado de mantener cada persona identificada debería realizarlo la cámara, con las consecuentes dificultades que eso conlleva cuando se producen oclusiones o cambios de luz y el consecuente aumento del tiempo de procesado. Además, en el caso de que la gente ande en grupos y sus pies se junten demasiado, la extracción de la información por parte del láser no será muy fiable ya que los datos se mezclarán y la posición de los pies se perderá.

2.1.3 Aproximación propuesta: Una cámara y un sensor de distancia

En este método se ha querido hacer uso de los elementos más habituales que se pueden encontrar ya desplegados en los lugares para los que está principalmente enfocado este proyecto, es decir, una cámara. En esta aproximación, se ha escogido como elemento extra un sensor de distancia transportado por el usuario ya que es

poco invasivo y que podría incluirse en lectores de códigos de barras para ver precios, carros de la compra, etc.

La arquitectura de esta aproximación puede encontrarse en la figura 1 de la presente memoria.

Con respecto al resto de aproximaciones, se tiene la principal ventaja de que el sensor de distancia es capaz de identificar cada objetivo de forma precisa, ya que se sabe qué sensor está dando cada medida y se reduciría la importancia que tiene el efecto de la oclusión en la obtención de la trayectoria. Puesto que la oclusión es un fenómeno bastante habitual en algunos de los lugares en los que se puede aplicar esta solución (centros comerciales o tiendas), representa una gran mejora.

También, como gran ventaja, este método es capaz de auto regularse e identificar la trayectoria seguida por cada persona a la hora de producirse cruces entre ellas sin hacer uso de ningún procesado en las cámaras, capacidad que no tienen las dos aproximaciones explicadas al principio de este capítulo. Como cada persona se encuentra identificada de forma inequívoca por un sensor de distancia, se sabría después del cruce qué trayectoria ha seguido cada persona. En el caso de las otras dos aproximaciones explicadas anteriormente, no se sabría qué trayectoria habría seguido cada persona (si han seguido recto o han girado).

2.2 Fases para solucionar el problema

2.2.1 Calibración

La calibración es un paso necesario en la visión por ordenador 3D para extraer información de mediciones de imágenes en 2D. En este capítulo se hará una revisión de los tipos de calibraciones que podemos encontrar [4] que se han usado por comunidades de fotogrametría [5] [6] y más recientemente, en la visión por ordenador [7] [8] [9].

Según el tamaño de los objetos que se usan para realizar la calibración, se puede clasificar estas técnicas en tres grandes grupos:

- Calibración basada en objetos 3D de referencia

La calibración de la cámara se realiza observando un objeto cuya geometría es de tres dimensiones. Este tipo de calibración es conocida por ser de gran precisión y de gran eficiencia [10].

En este Proyecto Final de Carrera, se ha escogido realizar esta calibración y encontrará una explicación detallada en el capítulo 3 de diseño.

- Calibración basada en planos 2D

Las técnicas usadas en esta categoría se desarrollan mediante la observación de un patrón 2D en diferentes orientaciones como el estudiado en [11, 12]. Debido a que casi cualquier persona puede crear un patrón de calibración por sí mismo gracias a su sencillez, este tipo de técnica de calibración es comúnmente elegida.

- Calibración basada en líneas de 1D

Los objetos usados para la calibración de este tipo de técnicas están compuestos de un conjunto de puntos definidos sobre una línea. La cámara se calibra observando el movimiento de esta línea sobre un punto fijo como por ejemplo, una cadena de bolas colgada del techo [13].

- Auto calibración (0D)

Las técnicas usadas en esta categoría no hacen uso de un objeto de calibración y se puede considerar como una aproximación de 0D porque solo se requieren correspondencias de puntos en la imagen. Este tipo de calibración intenta estimar directamente los parámetros internos y externos de la cámara resultando un problema matemático mucho más complicado [4].

Ningún método de calibración es el mejor de todos. Dependiendo de la situación en la que el usuario se encuentra y a la que se tiene que enfrentar.

2.2.2 Segmentación del fondo

En esta sección se realizará una introducción al trabajo que se está realizando en el campo de la extracción del fondo, la estructura general que presentan este tipo de algoritmos, cuales son los principales problemas que afectan al modelado del fondo y en qué escenarios puede ser preferible qué algoritmo [14].

Durante los últimos años se han desarrollado varios métodos para extraer el fondo teniendo cada uno sus propias características, fortalezas y debilidades. La

evaluación de cada uno de esos algoritmos ayuda a centrarse en los problemas que quedan por resolver.

Como la extracción del fondo tiene que enfrentarse a bastantes retos, se introducen elementos extra a la propia definición como tal de extracción del fondo. Se añade un procesado posterior que permite mejorar el rendimiento general del sistema, siendo por ejemplo de ayuda a la hora de adaptarse a posibles cambios graduales de la intensidad de la luz o etiquetado de cada objeto detectado.

En la siguiente figura se muestra el proceso típico a la hora de realizar la detección del fondo:

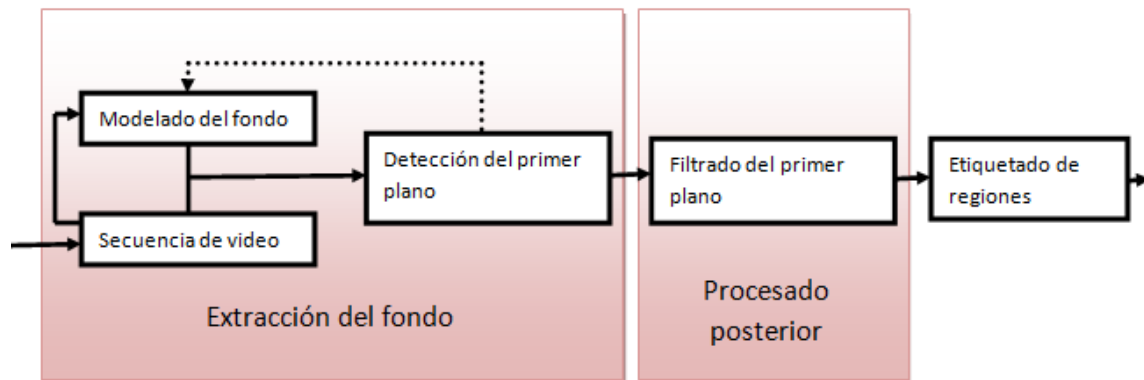


Figura 14. Esquema típico de un sistema de extracción de fondo

Los retos a los que generalmente tiene que enfrentarse un algoritmo de detección de fondo, en concreto en las dos primeras fases, son los siguientes:

➤ Cambios graduales de iluminación:

Es deseable que el modelado del fondo sea capaz de adaptarse a cambios graduales de la apariencia del ambiente. Así por ejemplo, en un escenario al aire libre, debería ser capaz de adaptarse al cambio gradual de la luz a lo largo del día.

➤ Cambios repentinos de iluminación:

Los cambios repentinos que se producen de vez en cuando no son cubiertos por un sistema de modelo de extracción del fondo. Estos cambios incluyen por ejemplo el encender unas luces en una habitación.

Para poder solventar este tipo de problemas, es necesaria una fase posterior que evalúe si esa mala detección es de interés o no.

➤ Fondo dinámico:

Algunos elementos que pertenecen al fondo podrían contener movimiento pero tendrían que ser tratados como fondo. Estos movimientos podrían ser periódicos o irregulares como la caída de las hojas, el movimiento de los árboles, las luces de los semáforos al controlar el tráfico, etc.

➤ Camuflaje

Intencionadamente o no, algunos objetos podrían ser difícilmente diferenciados del fondo haciendo la correcta clasificación más complicada. Como ejemplo, cuando una persona lleva la ropa del mismo color que las paredes que forman el fondo, no se sería capaz de diferenciar a qué pertenece ese pixel, si al fondo o al objeto en movimiento, en el caso de los métodos que hacen uso del color del pixel para detectar cuando hay un cambio de color en él y por tanto, algo que no pertenece al fondo.

➤ Sombras

Las sombras complican el procesado posterior en los pasos siguientes a la segmentación del fondo. Por ejemplo, podrían provocar que se juntaran detecciones de personas que estén bastante juntas y que pasaran a formar una sola detección, que generaría que se detectara esa detección como una persona o ninguna, dependiendo de cómo fueran sus proyecciones (fase de calibración comentada en el capítulo uno de la presente memoria).

➤ Ruido de la imagen

El ruido de la señal de video propio de la captura de la cámara degrada la imagen, que dificulta la correcta detección de los objetos en movimiento.

➤ Oclusión

Para poder realizar una correcta clasificación del objeto en movimiento en persona o no persona, es necesario que se tenga una silueta lo más clara posible para poder obtener sus proyecciones de forma precisa (ejemplo de este problema se puede encontrar en la figura 4 de la presente memoria).

En cuanto un objeto es ocluido su silueta puede variar de muchas formas e incluso dividir el objeto en movimiento en dos trozos y que para el sistema, contaría como dos objetos en movimiento en vez de uno ocluido.



Figura 15. Oclusión (imagen obtenida de <http://crcv.ucf.edu/projects/Part-basedTracking/index.php/>)

En la siguiente tabla se muestra un resumen de los métodos más relevantes y de qué característica principal hacen uso para el modelado del fondo [14]:

Método (primer autor)	Año	Característica
McFarlane	1995	Color
Stauffer	1999	Color
Oliver	2000	Correlación entre píxeles(color)
McKenna	2000	Croma, gradiente
Li	2003	Color
Kim	2004	Color, luminancia
Zivkovic	2006	Color
Maddalena	2008	Color
Barnich	2009	Color

Tabla 1. Métodos de modelado del fondo

2.2 3 Clasificación de los objetos detectados

Es muy importante reconocer de qué clase es un objeto detectado para poder hacer seguimiento de forma segura y poder analizar los datos después de forma correcta. Actualmente, hay dos principales grupos de técnicas que clasifican un objeto en movimiento. Un grupo de técnicas se basa en la forma física del objeto y otro grupo en el tipo de movimiento del objeto [15].

Las técnicas basadas en la forma física del objeto hacen uso de información en el espacio 2D del objeto. Las técnicas basadas en el movimiento del objeto hacen uso del seguimiento de un conjunto de características del movimiento del objeto.

Para el primer caso, la aproximación llevada a cabo en [16] hace uso de la longitud del contorno de la silueta y del área para clasificar objetos en diferentes grupos. Además, existe también la aproximación, que es la que se ha llevado a cabo en este proyecto, donde se obtienen unas proyecciones que identifican a las personas diferenciándolas del resto por el tipo de estructura que tiene su cuerpo [1].

También, para realizar un clasificador que sea capaz de diferenciar entre personas o no personas, se usan redes neuronales como clasificadores teniendo como entradas a la red características visuales de los objetos detectados. Los métodos estudiados en [17] [18] [19] hacen uso de este sistema.

En general, los métodos que hacen uso de características temporales del movimiento permiten distinguir objetos no rígidos como los humanos de objetos rígidos como vehículos, etc.

Para realizar este clasificado, la aproximación llevada a cabo en [20] se basa en la idea de que las personas tienen un movimiento al andar muy específico por su estructura y movimientos implicados (como los brazos) que las diferencian del resto. Además, se hace uso también de la idea de que el movimiento es muy parecido para todas las personas pero con pequeños cambios locales, lo que permite además identificar qué persona es de todas las detectadas hasta el momento. Esta aproximación crea un historial de apariencia de una persona mediante el cálculo de esos cambios locales y estructura general del movimiento. La clasificación final se realiza buscando repeticiones de estos movimientos y cambios locales permitiendo tanto distinguir si el objeto es persona o no y qué persona en concreto es.

El método propuesto en [21] se basa en una medida de auto similitud temporal del movimiento. Un objeto que exhibe evoluciones periódicas del movimiento, la medida de auto similitud también muestra un movimiento periódico.

2.2.4 Fusión de sensores

La integración de todos los sensores permite aumentar la precisión de la localización de los objetos y, por lo tanto, crear sistemas más robustos y con mejor rendimiento. Esta práctica es muy común en el uso de radares [22] y que, gracias a la

evolución de las CPU dando una capacidad de procesamiento mucho mayor, es posible aplicarse en la video vigilancia y campos relacionados.

La arquitectura general de los sistemas descritos en este proyecto sigue la siguiente línea [23]:

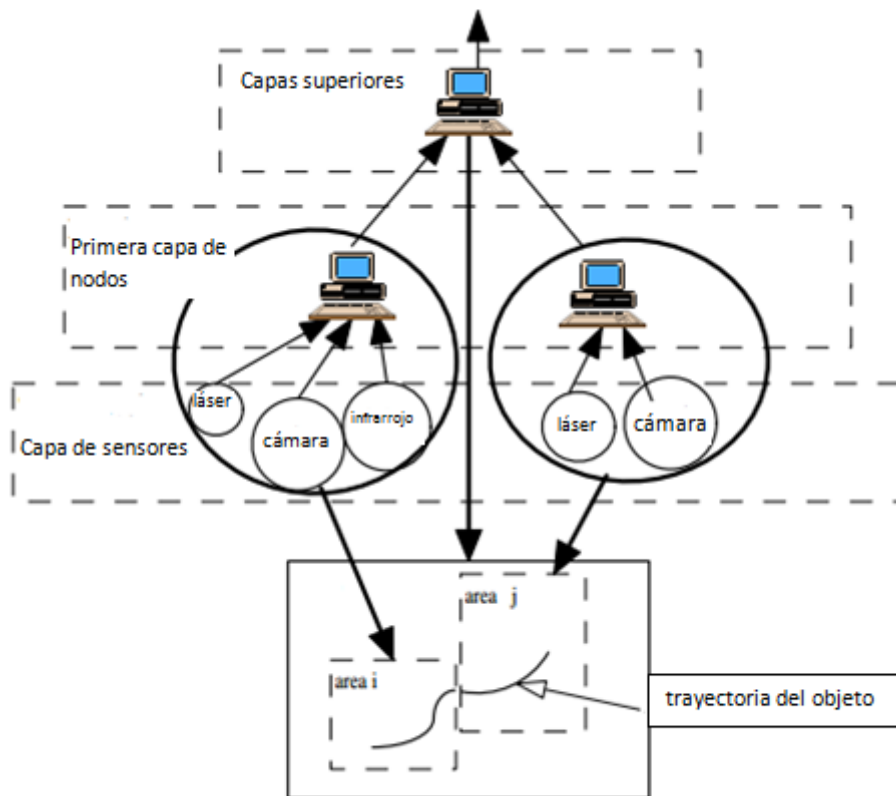


Figura 16. Esquema de un sistema de seguimiento

Los sensores que cubren la misma área se encuentran conectados a un nodo de procesamiento local común que es responsable de recolectar la información obtenida por los sensores que cubren esa zona. Cada nodo de procesamiento local se encarga de hacer también el procesado de imagen si los sensores no lo pueden hacer por sí mismos. Las capas superiores de nodos de procesamiento serán las encargadas de, basándose en la información que proporciona cada nodo de procesamiento local, calcular la trayectoria seguida a lo largo de todas las áreas.

En el caso de que el área a cubrir no sea lo suficientemente grande para tener varias capas de nodos y varios nodos locales, sería el nodo local de la primera capa el que haría todo el procesado y tarea de la fusión de los sensores que se conectan a él.

La fusión de todos los sensores es comúnmente realizada mediante el filtro de Kalman con el fin de obtener mejores estimaciones teniendo en cuenta todos los datos que aportan los sensores. El filtro de Kalman consiste principalmente en tres pasos: (1) obtener las mediciones disponibles de cada sensor como entrada; (2) creación de una estimación en base a los datos aportados como entrada y al estado del filtro; (3) actualización del estado del filtro basándose en la última estimación realizada.

Capítulo 3: DISEÑO

A lo largo de este capítulo se va a describir con detalle las técnicas implementadas para la realización de la detección de trayectorias. Incluiremos la secuencia de fases y su correspondiente descripción.

3.1 Perspectiva general

En la siguiente figura se muestra cuáles son las cuatro etapas principales que han sido necesarias analizar y realizar para conseguir el objetivo final y que se explicarán con detalle a lo largo de este capítulo.

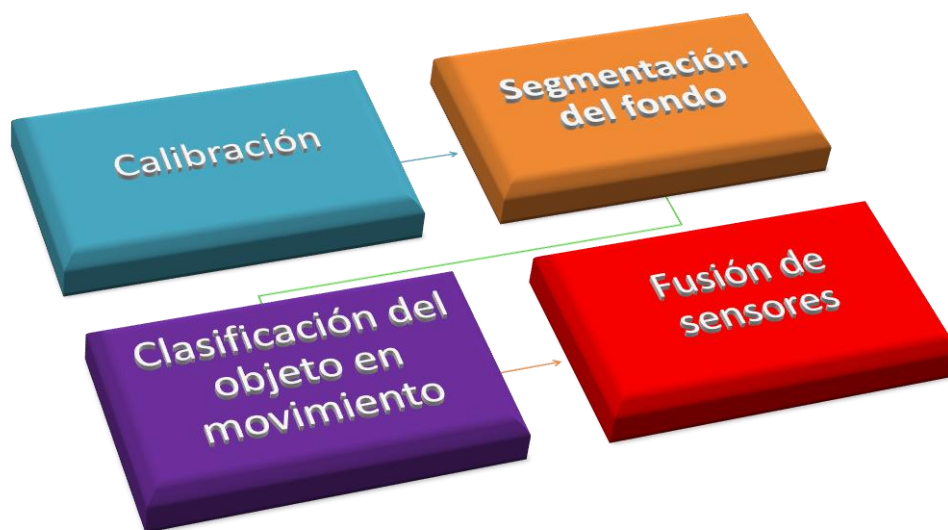


Figura 17. Etapas principales del proyecto

3.2 Calibración

A lo largo de esta sección se encontrará una explicación de qué es la calibración, qué nos permite obtener y qué pasos se han seguido para su realización [4].

La calibración es un proceso mediante el cual se obtiene la perspectiva de distancias en función de cómo está colocada la cámara ya que será necesario, al final,

poder obtener las tres coordenadas de la figura objetivo. Esta tarea es muy importante ya que permite adaptarse a cualquier escenario y a cualquier colocación de la cámara tanto en posición como en orientación sobre el eje en el que se coloca a la pared.

La fase de calibración calcula la matriz de proyección que relaciona un punto en el espacio 3D y el punto en la pantalla de la cámara, en el espacio 2D. Para poder realizar este cálculo, es necesario primero obtener algunas relaciones 3D -> 2D en el lugar de aplicación, llamados puntos de calibración.

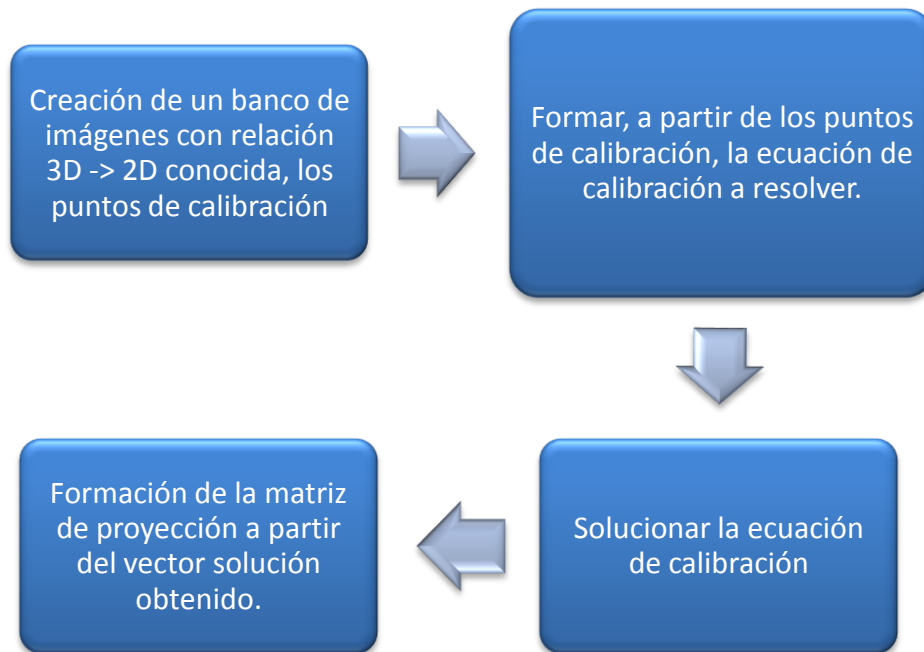


Figura 18. Diagrama de flujo de la sección de calibración

La matriz de proyección tiene la siguiente representación matricial:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (2)$$

La matriz de proyección a su vez puede dividirse en un vector de traslación y una matriz de rotación. Esta división de la matriz de proyección se representa en la figura siguiente:

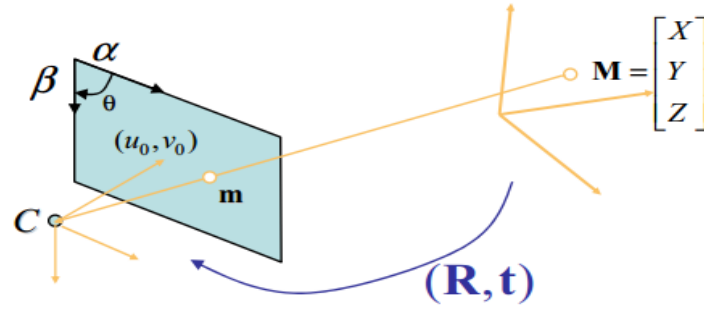


Figura 19. Representación del vector de traslación y matriz de rotación (imagen obtenida de [4])

El punto en el espacio 2D se denota como $m = [u, v]^T$. El punto en el espacio 3D se denota como $M = [X, Y, Z]^T$.

Es posible realizar un post procesamiento sobre el vector de traslación y la matriz de rotación para mejorar la precisión de cálculo de la matriz de proyección.

Para poder realizar la calibración es necesario saber al menos la relación del punto en 3D y su proyección en 2D de seis puntos. Son necesarios al menos seis puntos para poder evitar combinaciones lineales entre ellos y así, ser capaces de calcular una matriz de calibración que sea válida para toda la imagen. [24]

La ecuación que es necesario resolver se formula de la siguiente forma:

$$\begin{bmatrix} x^1 & y^1 & z^1 & 1 & 0 & 0 & 0 & 0 & -U^1x^1 - U^1y^1 - U^1z^1 \\ 0 & 0 & 0 & 0 & x^1 & y^1 & z^1 & 1 & -V^1x^1 - V^1y^1 - V^1z^1 \\ & & & & & & & \vdots \\ & & & & & & & \vdots \\ & & & & & & & \vdots \\ x^6 & y^6 & z^6 & 1 & 0 & 0 & 0 & 0 & -U^6x^6 - U^6y^6 - U^6z^6 \\ 0 & 0 & 0 & 0 & x^6 & y^6 & z^6 & 1 & -V^6x^6 - V^6y^6 - V^6z^6 \end{bmatrix} * \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix} = \begin{bmatrix} U^1 \\ V^1 \\ U^2 \\ V^2 \\ U^3 \\ V^3 \\ U^4 \\ V^4 \\ U^5 \\ V^5 \\ U^6 \\ V^6 \end{bmatrix} \quad (3)$$

X
 C
 Y

donde X es la matriz formada por los seis puntos con sus relaciones 3D \rightarrow 2D, el vector Y está formado por las coordenadas 2D de los seis puntos usados y el vector C es el vector a calcular y cuyas posiciones reordenadas forman la matriz de proyección añadiendo un valor extra, $P_{34} = 1$ que representa un factor de escalado [25].

La solución a esa ecuación se puede escribir de la siguiente manera:

$$C = (X^T X)^{-1} X^T Y = X^+ Y \quad (4)$$

donde la matriz X^+ es la matriz pseudo inversa de la matriz X .

Por tanto, una vez construida la matriz X y el vector Y usando los seis puntos anteriormente indicados y sus correspondencias entre el espacio 3D y el 2D, se resuelve la ecuación usando el comando “pinv” de MatLAB que calcula la pseudo inversa.

Para poder realizar nuestra investigación, se ha hecho uso de las siguientes imágenes de calibración:

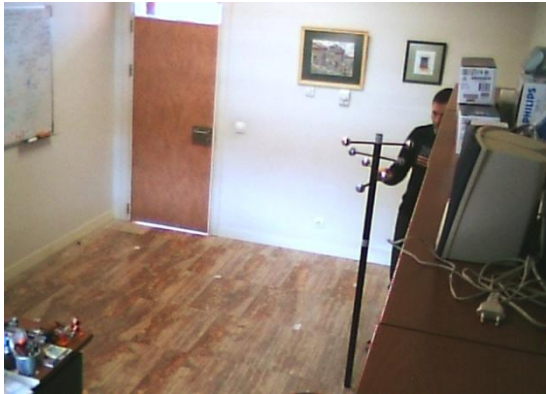


Figura 20. Imagen calibración 1

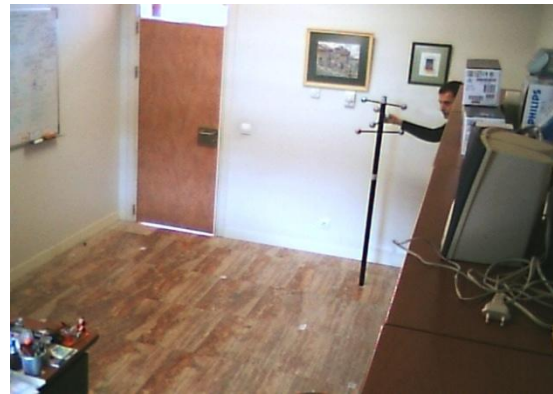


Figura 21. Imagen calibración 2



Figura 22. Imagen calibración 3

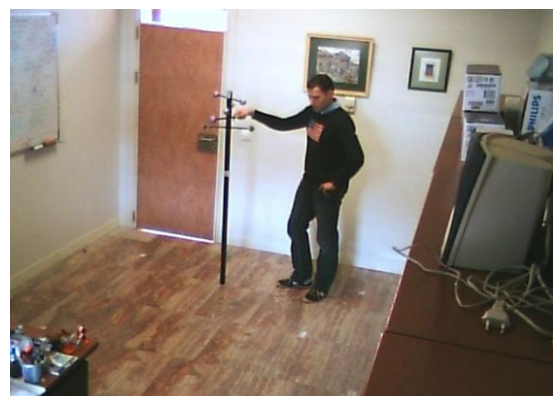


Figura 23. Imagen calibración 4

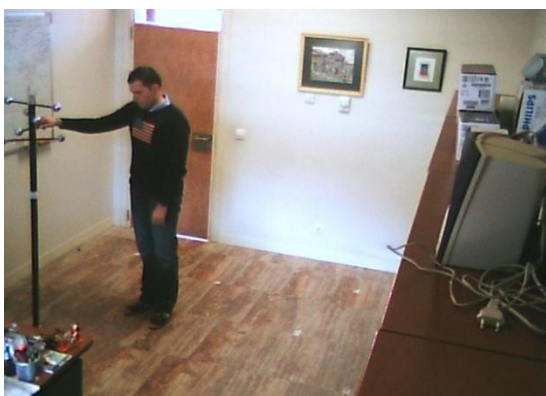


Figura 24. Imagen calibración 5

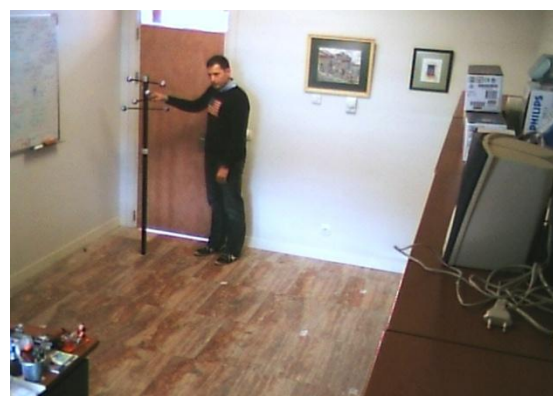


Figura 25. Imagen calibración 6

3.2 Segmentación del fondo

En esta sección encontrará una introducción de qué se obtiene en la tarea de extraer el fondo, de qué características se hace uso y qué algoritmos se prueban, cómo se puede adaptar a posibles cambios y qué proceso se ha seguido para realizar esta tarea satisfactoriamente.

La idea principal consiste en obtener una imagen de referencia del lugar de aplicación que se compara con cada frame que captura la cámara y detecta los píxeles que no pertenecen al fondo y por tanto, en movimiento.

La comparación de la imagen capturada por la cámara y la imagen de referencia del fondo se realiza pixel a pixel mediante dos métodos: uno evalúa la distorsión de la luz y de la croma (información sobre el color que hay en cada imagen) y otro que observa si los bordes de todos los objetos en la imagen han variado debido a que el objeto en movimiento hará cambios en los bordes.

Para mantener la actualización del fondo, una vez que se empieza la detección de movimiento, se mantiene entrenando al sistema para ser capaces de adaptarse a posibles cambios de fondo estáticos como movimientos de sillas, etc.

El flujo que sigue la detección de movimiento para llevar a cabo sus tareas satisfactoriamente es el siguiente:

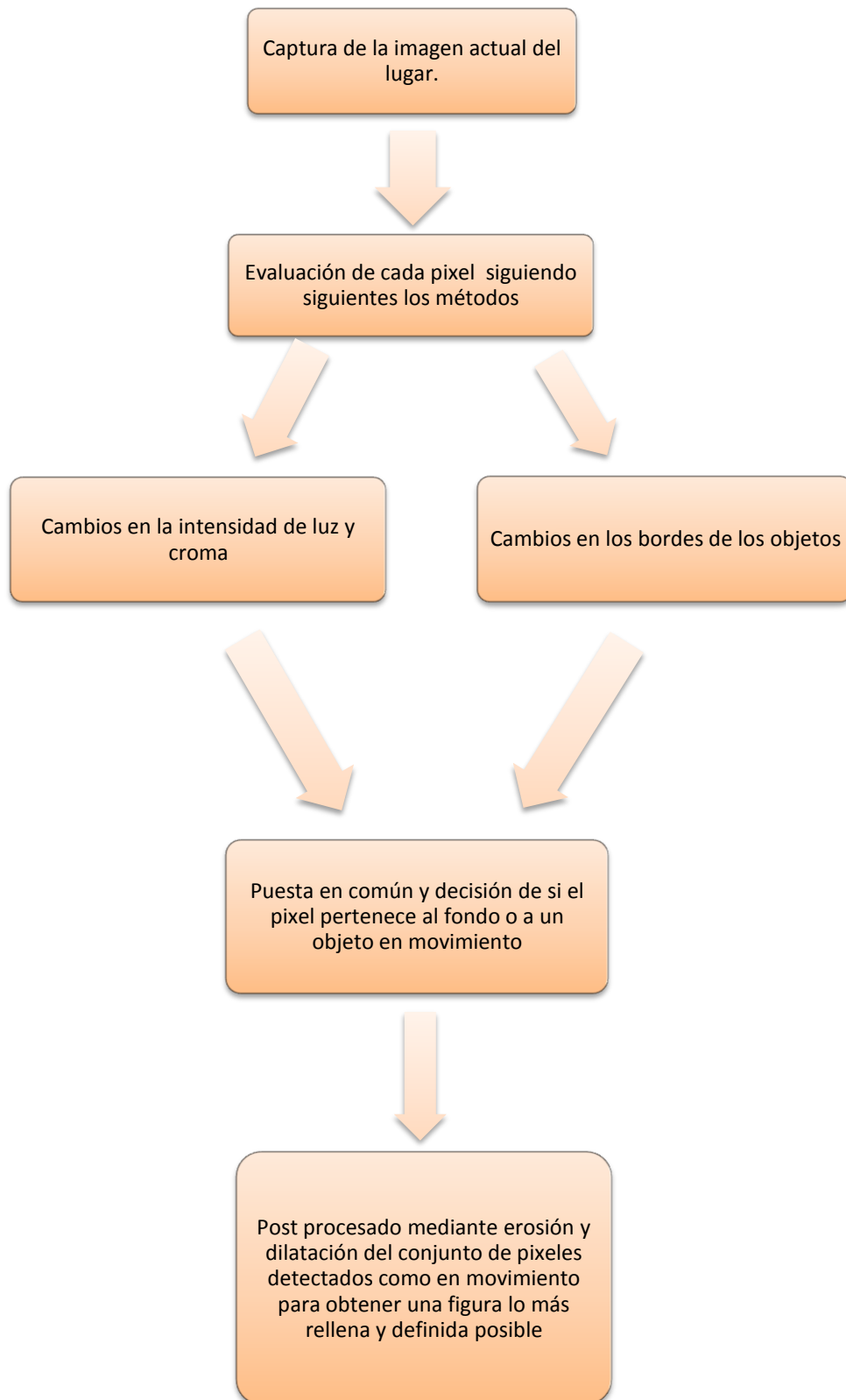


Figura 26. Diagrama de flujo de la extracción del fondo

Se pueden encontrar múltiples algoritmos para obtener el fondo. En este proyecto se han realizado las siguientes aproximaciones:

- i. Mixture of Gaussians background extraction
- ii. Illuminance and chrominance extraction
- iii. Edge background extraction

3.2.1 Mixture of Gaussians background extraction

En este método se tiene en cuenta la serie de valores de los píxeles en el espacio de iluminancia y croma. [26]

La observación del píxel “ i ” en el tiempo “ t ” se representa como

$$X_{i,t} = [Y_{i,t} \quad U_{i,t} \quad V_{i,t} \quad D_{i,t}]$$

El historial de observaciones $[X_{i,1}, \dots, X_{i,t-1}]$ se representa como una mezcla de K distribuciones gaussianas. Típicamente K varía entre 3 y 5 dependiendo de la capacidad computacional.

Por tanto, la probabilidad de que en la observación actual, el píxel i pertenezca a una de las gaussianas siguiendo el modelo indicado en el párrafo anterior, se podría estimar como:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (5)$$

donde η es la función de densidad probabilidad Gaussiana, donde $\omega_{i,t-1,k}$ es el peso asociado a la Gaussiana K -ésima en el instante $t-1$ y donde $\mu_{i,t-1,k}$ y $\Sigma_{i,t-1,k}$ son la matriz media del espacio YUV D y la matriz de covarianza la gaussiana K -ésima.

Los pesos indicados anteriormente representan qué proporción de las observaciones pasadas son modeladas por cada gaussiana. Para actualizar a cada paso la mezcla de gaussianas, cuando se recibe una nueva observación se comprueba si alguna de las gaussianas corresponde con el valor observado.

Se considera que una observación corresponde a alguna de las gaussianas si la observación normalizada en función de la media y la varianza de esa gaussiana es menor de 2.5 de la desviación estándar.

Si alguna de las gaussianas coincide, su valor de media y varianza se actualiza de la siguiente forma:

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (6)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \quad (7)$$

y se mantiene igual para el resto de gaussianas, las que no han coincidido.

Hay que tener en cuenta que las varianzas no se dejan caer por debajo de un valor mínimo para que las correspondencias no se conviertan inestables en escenarios que son estáticos por periodos de tiempo demasiado largos. El parámetro ρ podría interpretarse como la tasa de aprendizaje.

Los pesos de las gaussianas se actualizan de la siguiente forma:

$$\omega_{i,t,k} = (1 - \alpha)\omega_{i,t-1,k} + \alpha M_{i,t,k} \quad (8)$$

donde $M_{i,t,k}$ es 1 para la gaussiana que ha dado correspondencia y 0 para el resto, lo que va produciendo que la gaussiana que corresponde al valor del pixel aumente su peso sobre ese pixel y el resto que no han correspondido baje su peso sobre ese pixel.

Hay que tener en cuenta que debido a los cambios de intensidad luminosa que puedan producirse, los valores de croma por debajo de un cierto valor de luminancia son inestables y por tanto habría que descartarlos a la hora de calcular los pasos anteriores.

Finalmente, para hacer el algoritmo más robusto, se calcula la actividad con la siguiente fórmula:

$$A_{i,t,k} = (1 - \lambda)A_{i,t-1,k} + \lambda |Y_{i,t} - Y_{i,t-1}| \quad (9)$$

$A_{i,t,k}$ es el valor de actividad del pixel i en el tiempo $t-1$ para la distribución gaussiana k y la matriz Y representa el valor de luminancia en el tiempo t y $t-1$. El parámetro λ se conoce por ser la tasa de aprendizaje.

Cuando la actividad supera un cierto umbral, la tasa de aprendizaje se reduce drásticamente para que la caracterización del fondo no se vea afectada por movimientos esporádicos de objetos.

El cálculo de la Actividad va a ser muy útil en escenarios donde ha de entrenarse el algoritmo con cierta actividad por cuestiones físicas, o simplemente debido a que el algoritmo debería funcionar incluso una vez entrenado y puesto en funcionamiento, ya que sería necesario adaptarse a cambios del fondo (nuevos elementos introducidos en el escenario que pasan a formar parte del fondo porque no se mueven)

Por último, se tiene que decidir qué gaussianas hay que mirar para cada pixel para ver si ese pixel corresponde al fondo o corresponde a un objeto en movimiento. Para ello, primero, una vez entrenado el algoritmo, se ordenan para cada pixel las gaussianas en orden descendente de los pesos normalizados por sus varianzas, ω/σ ,

ya que interesan las gaussianas con mucho peso sobre ese pixel y que la varianza sea pequeña para que se pueda tener una detección lo más exacta posible.

Después, se calculará el número de gaussianas mínimas que son necesarias para caracterizar el fondo de cada pixel mediante la siguiente fórmula:

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b w_k > T \right) \quad (10)$$

que representa cuantas gaussianas son necesarias para obtener una proporción T del pixel.

En la siguiente imagen se representa como una de las gaussianas ha obtenido la mayor importancia sobre un pixel y representa la mayoría de los valores que ha tenido ese pixel durante la fase de entrenamiento.

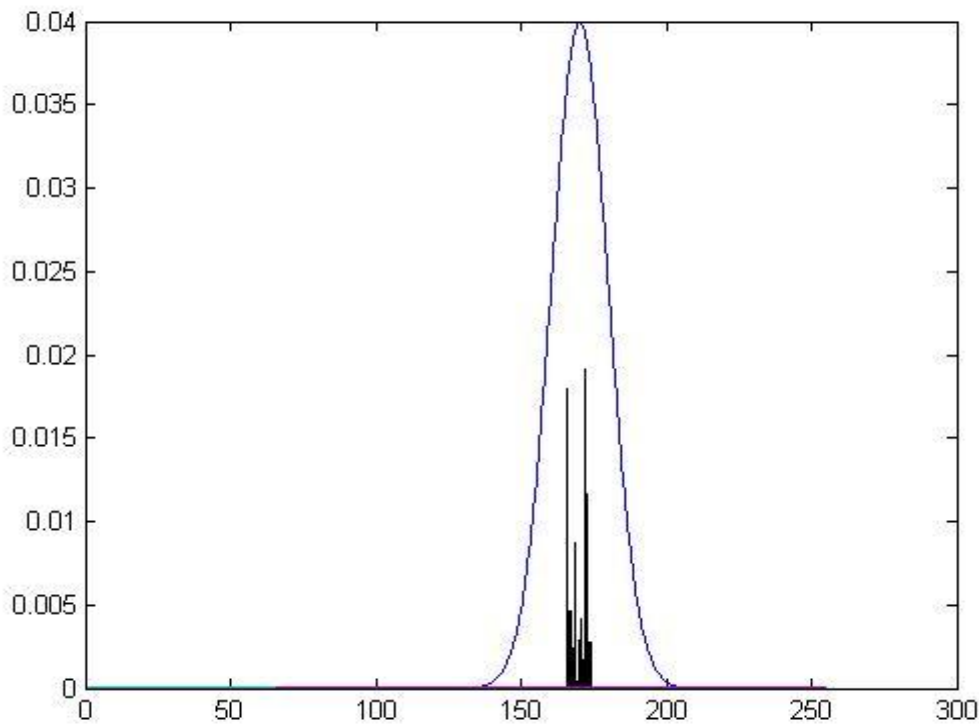


Figura 27. Peso de las gaussianas sobre un pixel

El resto de gaussianas no tienen ningún peso sobre ese pixel y por eso aparecen con valor cero.

En este caso, para comprobar si el valor actual de este pixel se encuentra en el fondo o no, tendremos que normalizar su valor con la media y desviación estándar de esta gaussiana y se comprobaría si tiene una desviación estándar entre ± 2.5 .

3.2.2 Luminance and chrominance distortion background extraction

La principal idea en la que se basa este método es comparar la imagen actual con la imagen de referencia de fondo estático extraída a lo largo de N frames separando por un lado el brillo y por otro la croma [27]. Obteniendo para cada iteración la distorsión que sufren ambos componentes permitirá saber cuando un elemento pertenece al fondo o no.

La siguiente figura ilustra estos componentes en el espacio 3D.

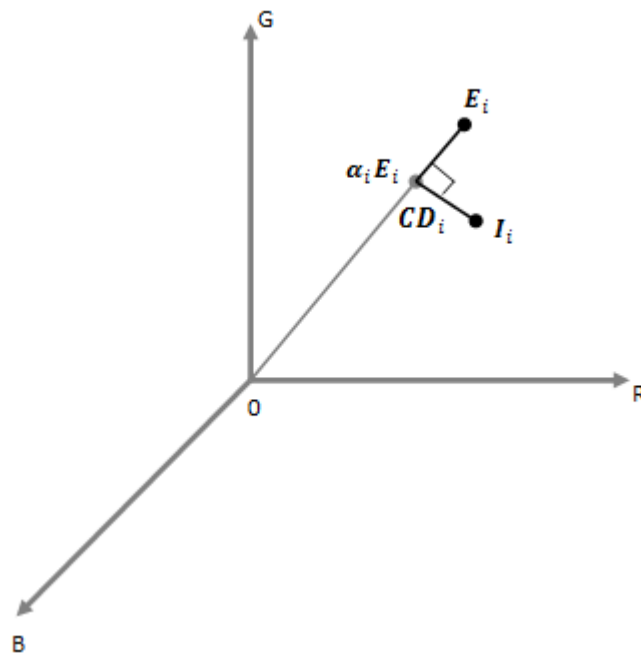


Figura 28. Separación de la luz en sus componentes de luminancia y croma

✓ La distorsión del brillo

Es un escalar que representa la distancia entre el color observado y la línea de croma esperada. Se obtiene de minimizar la siguiente fórmula:

α_i representa la fuerza de brillo con respecto al valor esperado. Si el valor del α_i es 1, el brillo del pixel observado es el mismo que el de referencia. Si es menor que 1, tiene menos brillo, si es mayor que 1, tiene más brillo.

$$\phi(\alpha_i) = (I_i - \alpha_i E_i)^2 \quad (11)$$

✓ *Distorsión del color*

La distorsión del color está definida como la distancia ortogonal entre el color observado y la línea de croma esperada. La distorsión de color está definida como:

$$CD_i = || I_i - \alpha_i E_i || \quad (12)$$

Este algoritmo se divide en tres fases:

- i. Modelado del fondo
- ii. Selección de umbrales
- iii. Clasificación de pixeles

i. Modelado del fondo

En esta fase, cada pixel estará formado de la siguiente forma:

$$< E_i, s_i, a_i, b_i >$$

donde E_i es el valor de color esperado del pixel i que se describe con la siguiente fórmula:

$$E_i = [\mu_R(i), \mu_G(i), \mu_B(i)]$$

donde $\mu_R(i), \mu_G(i), \mu_B(i)$ son las medias aritméticas de los valores rojo, verde y azul del pixel respectivamente a lo largo de N frames.

S_i se define como la desviación estándar y se representa de la siguiente forma:

$$s_i = [\sigma_R(i), \sigma_G(i), \sigma_B(i)]$$

Para normalizar las bandas de color, las ecuaciones 11 y 12 se convierten en:

$$\alpha_i = \min[\sum_{C=R,G,B} \left(\frac{I_C(i) - \alpha_i \mu_C(i)}{\sigma_C(i)} \right)^2] = \frac{\left(\frac{I_R(i)\mu_R(i)}{\sigma_R^2(i)} + \frac{I_G(i)\mu_G(i)}{\sigma_G^2(i)} + \frac{I_B(i)\mu_B(i)}{\sigma_B^2(i)} \right)}{\left(\left[\frac{\mu_R(i)}{\sigma_R(i)} \right]^2 + \left[\frac{\mu_G(i)}{\sigma_G(i)} \right]^2 + \left[\frac{\mu_B(i)}{\sigma_B(i)} \right]^2 \right)} \quad (13)$$

$$CD_i = \sqrt{\sum_{C=R,G,B} \left(\frac{I_C(i) - \alpha_i \mu_C(i)}{\sigma_C(i)} \right)^2} \quad (14)$$

Diferentes pixeles producen distribuciones diferentes de α y CD. Para hacer el análisis de los datos, se necesita comparar las variaciones entre los componentes de brillo y croma. Por lo tanto, se calcula la distancia entre α y αE_i y su RMS (Root Mean Square) o error cuadrático medio.

$$\alpha_i = RMS(\alpha_i) = \sqrt{\frac{\sum_{i=0}^N (\alpha_i - 1)^2}{N}} \quad (15)$$

$$b_i = RMS(CD_i) = \sqrt{\frac{\sum_{i=0}^N (CD_i)^2}{N}} \quad (16)$$

a_i es la variación de la distorsión del brillo y b_i es la variación de la distorsión de la croma.

ii. Selección de umbrales

Para realizar la selección de umbrales, se realiza un banco de pruebas con diferentes configuraciones de umbrales para calcular el número de clasificaciones erróneas que se obtienen.

Para poder hacer estos umbrales lo más escalables posibles, es necesario probar en diferentes escenarios y bajo diferentes condiciones usando los mismos umbrales en todas esas configuraciones para llegar a unos valores que en promedio den el mejor resultado.

iii. Clasificación de los pixeles

En esta tercera fase, se va a evaluar la diferencia entre la imagen de fondo y la imagen actual y así ser capaces de decidir que pixel pertenece al fondo y que pixel pertenece al objeto en movimiento.

Es necesario comparar los valores obtenidos de brillo y croma con unos umbrales para poder decidir qué tipo de pixel es. Estos umbrales han sido decididos en la fase anterior haciendo uso de esta fase, que nos explica qué valores hay que comparar con qué umbral y de la forma que hay que hacerlo (mayor que, menor que, etc.)

A continuación se presenta una tabla donde se muestra la correspondencia de cada pixel en función de sus umbrales:

$$M(i) = \begin{cases} F : & \widehat{CD}_i > \tau_{CD} \text{ or } \widehat{\alpha}_i < \tau_{\alpha low}, & else \\ B : & \widehat{\alpha}_i < \tau_{\alpha 1} \text{ and } \widehat{\alpha}_i > \tau_{\alpha 2}, & else \\ S : & \widehat{\alpha}_i < 0, & else \\ H : & & \end{cases}$$

Tabla 2. Clasificación de los pixeles

F: Objeto en movimiento -> el pixel tiene diferente croma del original

B: Perteneciente al fondo -> el pixel tiene croma y brillo similar.

S: Fondo sombreado o sombra -> el pixel tiene croma similar pero brillo más bajo.

H: fondo destacado -> croma similar y brillo más alto.

Como se ha indicado anteriormente, cada pixel tiene una distribución diferente de α y CD y para poder usar un umbral único, es necesario, para cada pixel, re escalar ambos valores de la siguiente forma:

$$\hat{\alpha}_i = \frac{\alpha_i - 1}{a_i} \quad (17)$$

$$\widehat{CD}_i = \frac{CD_i}{b_i} \quad (18)$$

En las siguientes imágenes se puede comprobar la variación de la distorsión del brillo y de la distorsión de la croma con un entrenamiento de 2400 frames.



Figura 29. Variación de la distorsión del brillo



Figura 30. Variación de la distorsión de la croma

3.2.3 Edge background extraction

Este algoritmo de segmentación del fondo se basa en la variación de la magnitud y orientación de los bordes de los objetos de la imagen entre dicha imagen y una imagen de referencia obtenida previamente en una fase de entrenamiento. [28]

Para cada matriz de color (rojo, azul y verde) que forma la imagen se tiene que obtener la diferencia entre la imagen actual y la de referencia (imagen media) en el eje X y en el eje Y . Dicho cálculo se realiza de la siguiente forma:

$$\Delta H = |H - H_t|, \Delta V = |V - V_t| \quad (19)$$

donde H y V son los bordes horizontales y verticales de la imagen actual y H_t y V_t representan la media de los bordes horizontales y verticales y que forman la imagen de referencia.

A continuación, se define un factor de confianza, llamado R , como el % de la magnitud del borde que ha cambiado.

Si por ejemplo, un borde tiene una magnitud de 100 y otro de 50 y ambos cambian una cantidad de 10, uno de ellos habrá cambiado un 10% y otro lo habrá hecho en un 20%. Esta variación es la que se mide con el factor de confianza, R .

Para calcularlo, vamos a usar la siguiente fórmula:

$$R = \frac{\Delta G}{G_t^*} \quad (20)$$

donde $\Delta G = \Delta H + \Delta V$ y se define como el gradiente de la imagen y

$$G = |H| + |V|, \quad G_t = |H_t| + |V_t|, \quad G_t^* = \max\{G, G_t\} \quad (21)$$

Una vez que este procesado se ha hecho, se calculará la confianza de los píxeles de la siguiente manera:

$$C^e = \begin{cases} 0\% & R\Delta G < m\sigma \\ \frac{R\Delta G - m\sigma}{M\sigma - m\sigma} \times 100\% & m\sigma \leq R\Delta G \leq M\sigma \\ 100\% & R\Delta G > M\sigma \end{cases}$$

Tabla 3. Cálculo de la confianza para un píxel

donde σ es la suma de la desviación estándar en la dirección horizontal y vertical. M y m son constantes que han de ser calculadas mediante un banco de pruebas en el que se seleccionen los valores que menor error dan en la clasificación.

A continuación se presentan dos ilustraciones donde se pueden comprobar las imágenes de referencia obtenidas en la fase de entrenamiento para los bordes horizontales y verticales.

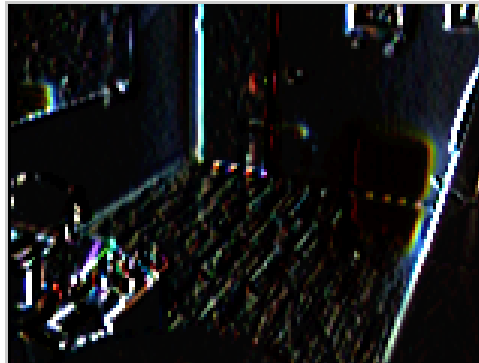


Figura 31. Imagen de referencia para los bordes verticales

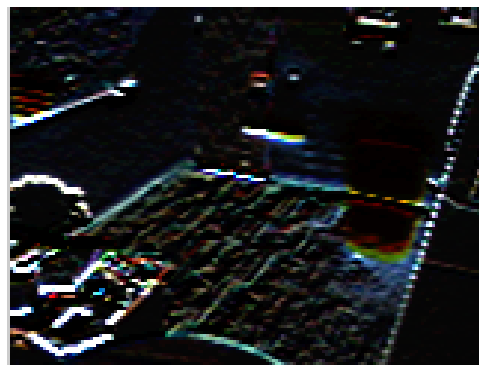


Figura 32. Imagen de referencia para los bordes horizontales

3.2.4 Procesado posterior de la imagen

Para acabar de obtener una captura de la figura en movimiento de la imagen de la cámara lo más nítida posible, se realiza un procesado mediante el algoritmo Connected Components (Componentes conectados) que elimina los falsos positivos, un filtrado usando el filtro de mediana que junta zonas de pixeles detectados muy densas y finalmente, un procesado morfológico basado en erosión y dilatación.

3.2.4.1 Connected Components

El algoritmo Connected Components nos permite etiquetar conjuntos de pixeles que se encuentran conectados entre sí.

En la teoría de grafos, un componente conectado (o simplemente componente) de un grafo es un sub grafo en el que cualquiera de los dos vértices están conectados entre sí por caminos.

Por ejemplo, el gráfico que se muestra en la siguiente imagen tiene tres componentes conectados.

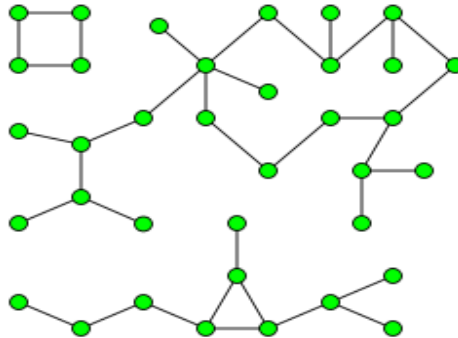


Figura 33. Grafo de representa tres caminos conectados (imagen obtenida de http://en.wikipedia.org/wiki/Connected_component_%28graph_theory%29)

Un gráfico que está conectado en sí tiene exactamente un componente conectado, que consiste en todo el gráfico.

3.2.4.2 Filtro de mediana

El filtro de mediana es un filtro digital en el dominio del espacio y no lineal.

Este filtro es comúnmente usado como pre procesamiento de imágenes para eliminar el ruido y mejora algunos procesados posteriores como la obtención de bordes.

En las siguientes figuras se puede comprobar alguno de los usos típicos del filtro de mediana y como mejora la calidad de la imagen. [29]

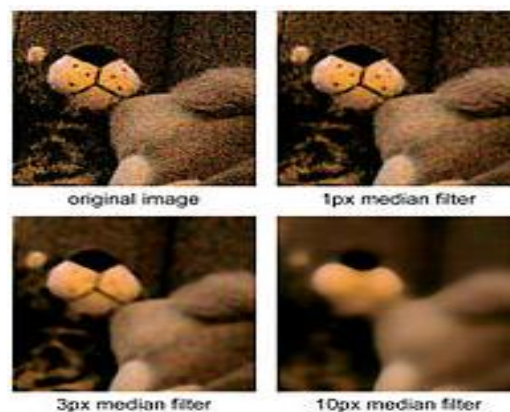


Figura 34. Tipos de filtros de mediana y su resultado (imagen obtenida de http://en.wikipedia.org/wiki/Median_filter)



Figura 35. Uso del filtro de mediana para corregir una imagen corrompida por la mala recepción (imagen obtenida de http://en.wikipedia.org/wiki/Median_filter)

La nueva imagen $g(x, y)$ se genera a base de hallar la mediana del conjunto formado por los píxeles de la imagen f , en un entorno de vecindad del punto (x, y) , repetidos tantas veces como se indique en la máscara $h(u, v)$.

Una máscara $h(u, v)$ comúnmente usada es mostrada en la siguiente ilustración:

1	2	1
2	4	2
1	2	1

Figura 36. Máscara

3.2.4.3 Erosión y dilatación

La erosión y la dilatación son las dos operaciones fundamentales en el procesamiento de imágenes morfológico.

De entre todas sus aplicaciones, en este proyecto se ha centrado en la parte de extracción de la frontera y rellenado de regiones para conseguir la silueta lo más definida posible y así poder identificar la figura como una persona o no de manera mucho más precisa.

La frontera de un conjunto A se puede obtener primero erosionando A por un elemento estructural apropiado, B , y realizando posteriormente la diferencia entre A y su erosión.

El elemento estructural B usado más frecuentemente es el cuadrado 3×3 (como en el ejemplo que se muestra a continuación). Usando otros tamaños, por ejemplo 5×5 , se ampliaría el grosor de la frontera a dos o tres píxeles.

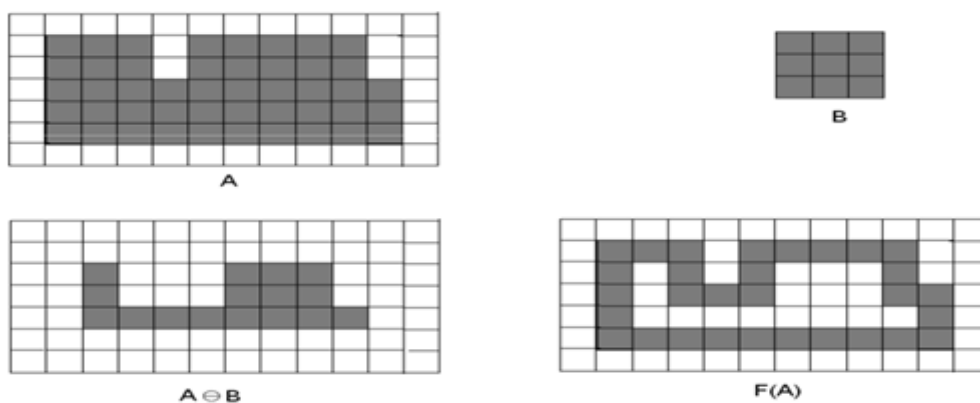


Figura 37. Extracción de la frontera mediante el uso de la erosión

Para el caso de la dilatación, se consigue rellenar una estructura basándonos en un elemento estructural previamente definido.

Partiendo del borde 8-conexo de una región, A , y de un punto p del interior de A . El siguiente procedimiento rellena el interior de A :

$$X_0 = p$$

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k = 1, 2, 3...$$

donde B es el elemento estructural siguiente:

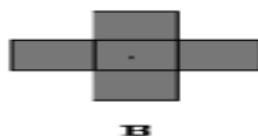


Figura 38. Elemento estructural usado en la dilatación

Y el algoritmo termina en la iteración k si $X_k = X_{k-1}$. La unión de X_k y A es la frontera y la región rellena. [30]

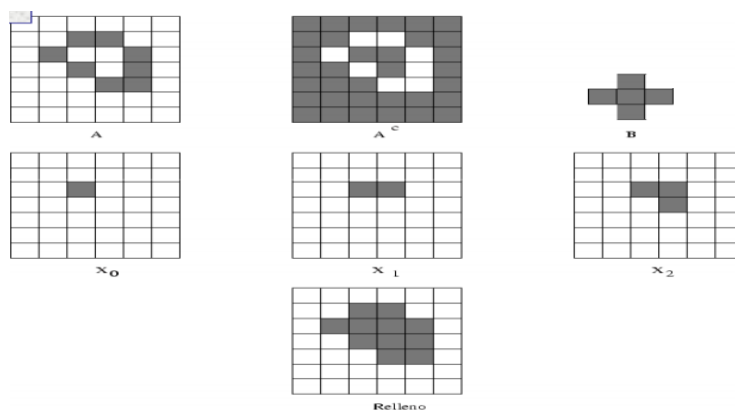


Figura 39. Uso de la dilatación para relleno de figuras

A continuación comprobaremos cómo ha dado resultado el post procesado.

- ✓ La imagen 40 es la imagen capturada con la cámara
- ✓ La imagen 41 es la imagen calculada solo en la fase de detección y solo con un post procesado de Connected Components
- ✓ La imagen 42 es la imagen calculada en la fase de detección con un post procesado de Connected Components, erosión y dilatación



Figura 40. Imagen capturada



Figura 41. Obtención del movimiento con Connected Components



Figura 42. Obtención del movimiento con todo el post procesado.

3.3 Clasificación de personas

En esta sección, se puede encontrar porqué es necesario realizar la clasificación, qué características de las personas se han elegido para intentar diferenciarlas del resto de objetos, qué pasos se han seguido para conseguir el objetivo final y qué algoritmo se ha usado para llegar al objetivo final de esta parte.

Una vez se ha realizado la detección del objeto en movimiento, es necesario saber si ese objeto es una persona o no puesto que solo en el primer caso se estaría interesado en continuar con el procesado. Además, se necesita saber qué punto de la imagen hay que usar como referencia a la hora de calcular sus coordenadas 3D una vez se ha confirmado si es persona o no.

Para poder realizar esta parte, se ha escogido hacer un estudio morfológico de la figura, ya que las personas tienen una forma peculiar, y sería suficiente para permitir los objetivos entre personas y el resto de objetos.

Este estudio morfológico llevado a cabo se basa en la obtención de las proyecciones verticales y horizontales del objeto en cuestión y comparar esas proyecciones con una base de datos previamente formada de proyecciones horizontales y verticales de objetivos que se sabía con certeza que eran personas. Este tipo de estudio de las personas permite incluso clasificar a esas personas a su vez según su postura [1].

Para conseguir este objetivo se han seguido los siguientes pasos:

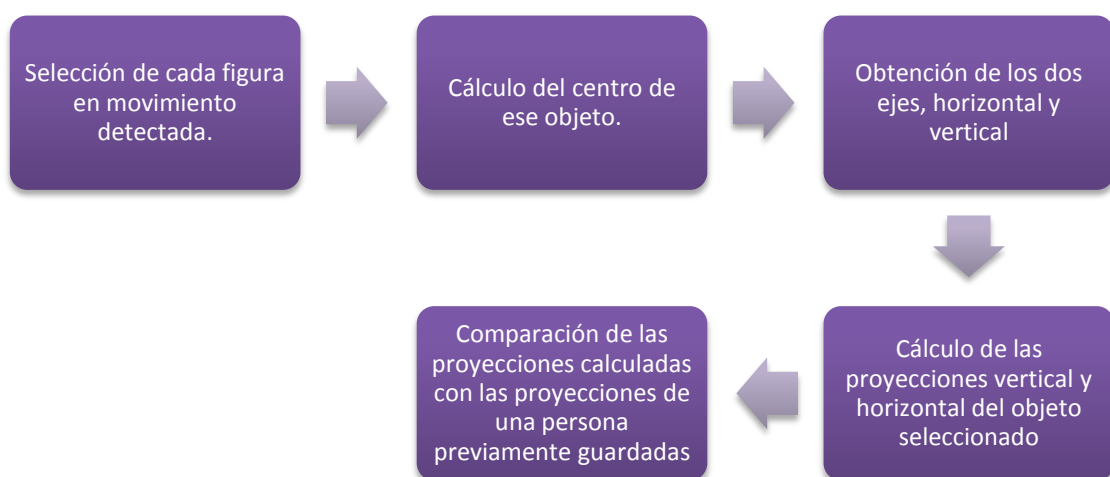


Figura 43. Diagrama de flujo de la clasificación de personas

Para obtener las proyecciones de la figura identificada y etiquetada será necesario realizar PCA (Principal Component Analysis).

Principal Component Analysis es una técnica usada en estadística para hacer reducción de dimensiones en un conjunto de datos de forma que sea más fácil analizarlos, clasificarlos, etc. y cuya gran ventaja es que retiene las características del conjunto de datos en cuestión que más contribuyen a su varianza y que generalmente son los de más interés.

PCA realiza una transformación lineal donde escoge un nuevo sistema de coordenadas para el conjunto de datos iniciales en los cuales la mayor varianza recae sobre el primer eje, la segunda mayor varianza sobre el segundo eje y así sucesivamente.

Para construir esta transformación lineal primero se construye una matriz de covarianza y después se calculan los auto valores y auto vectores.

✓ Pasos necesarios

1. Etiquetar objetos en movimiento de la imagen

Primero, es necesario etiquetar todos los objetos detectados en la imagen para poder evaluar cada uno individualmente. Para poder efectuar esta acción, se ha hecho uso de la función “regionprops” de MatLAB donde etiqueta cada objeto independiente de la imagen con un identificador y a su vez permite obtener varias características importantes como su centro y la longitud de los ejes mayor y menor.

2. PCA: Obtención del centro y del eje mayor y menor.

Para obtener el centro el centro de la figura y sus ejes se hace uso de la función “princomp” de MatLAB que calcula un conjunto de características sobre el objeto etiquetado entre las que se encuentran el centro y su eje mayor.

El eje más largo es el eje donde mayor varianza hay y por tanto una de las proyecciones necesarias para nuestro estudio. Una vez que se ha obtenido el eje más largo, es necesario obtener el eje para la segunda proyección y que se corresponde con el eje perpendicular al obtenido anteriormente.

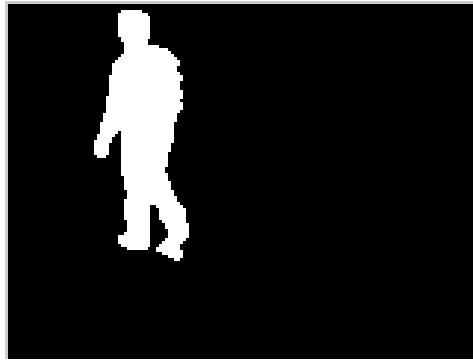


Figura 44. Imagen de referencia



Figura 45. Obtención del centro de la figura objetivo

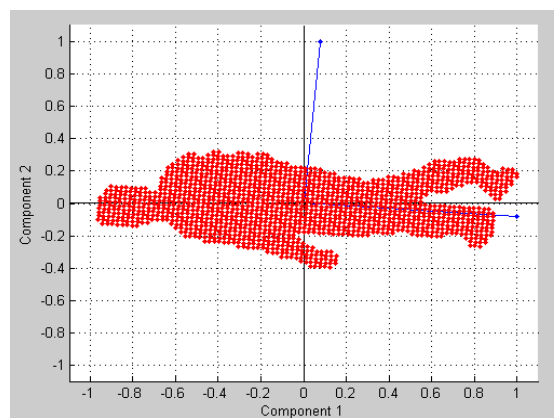


Figura 46. Obtención del eje de mayor varianza y su perpendicular (PCA)

3. Construcción de las proyecciones

Para construir las proyecciones horizontales es necesario realizar una proyección de todos los puntos que pertenecen a la figura detectada sobre ambos ejes y realizar un contador que te calcula cuantos puntos se proyectan en cada posición de cada eje y por tanto, al final, tener las proyecciones que se muestran a continuación.

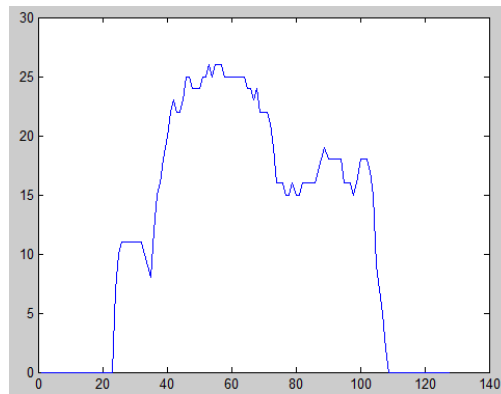


Figura 47. Proyección horizontal

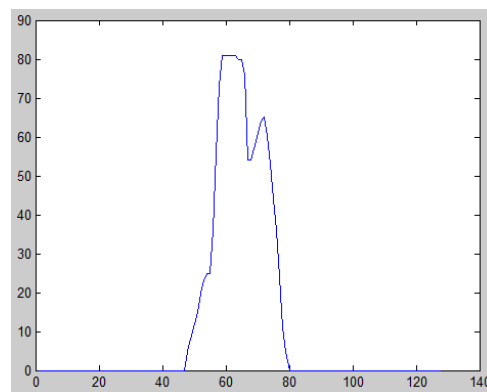


Figura 48. Proyección vertical

4. Toma de decisión

En este último paso, es necesario inicialmente crear una proyección horizontal y una vertical de referencia.

Para crear estas proyecciones de referencia se ha realizado la media aritmética de las proyecciones calculadas a partir de detecciones conocidas de personas.

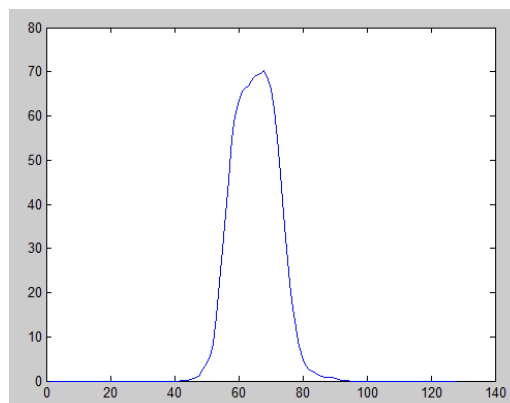


Figura 49. Proyección vertical de referencia

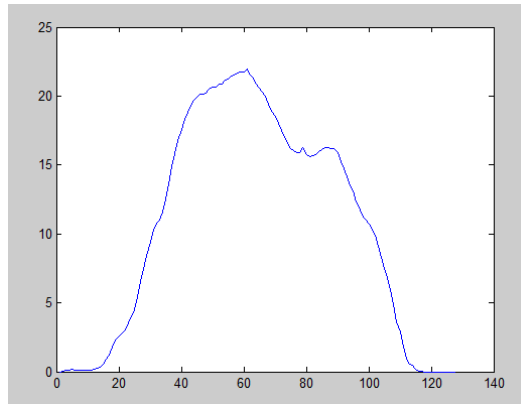


Figura 50. Proyección horizontal de referencia

Una vez que se tienen esas proyecciones de referencia, se calculará cuanto se parecen tanto la proyección horizontal como la vertical con las proyecciones vertical y horizontal de referencia mediante la siguiente fórmula:

$$S = \sum_{i=1}^{180} (H - H_t)^2 + (V - V_t)^2 \quad (22)$$

donde H y V hacen referencia a las proyecciones horizontal y vertical de la imagen a evaluar respectivamente y H_t y V_t son las proyecciones horizontal y vertical de referencia previamente calculadas.

En este proyecto, se han guardado las proyecciones en vectores de ciento ochenta posiciones y de ahí ese valor en el sumatorio de la formula anterior.

Para hacer una comparación correcta, es necesario centrar tanto la proyección horizontal como la vertical con respecto al centro de la figura. Según este criterio, la proyección del centro de la figura deberá caer sobre la posición noventa de cada proyección al tener estas 180 posiciones de longitud.

3.4 Fusión de sensores

En esta última fase, se consigue obtener las tres coordenadas del centro del objeto en movimiento detectado a lo largo de las fases anteriores. En esta fase es cuando entra en acción el sensor de movimiento que lleva consigo cada objetivo. Habrá que establecer el sistema de tres ecuaciones con tres incógnitas y resolverlo.

En este último apartado, además, se realizará una estimación de la trayectoria basada en el filtro de Kalman para corregir posibles errores producidos en la detección y que se basa en las posiciones anteriores calculadas y parámetros establecidos como la velocidad a la que se mueve una persona para estimar la siguiente posición.

Se simulará el comportamiento de un sensor de distancia que tenga la persona objetivo y por tanto se podrá saber en todo momento la distancia desde la persona al sensor. Con este dispositivo se obtendrá una ecuación extra, la de la distancia, en función de tres variables, las tres coordenadas en el espacio 3D de la persona.

A continuación, se muestra el orden de los pasos que se han seguido para solventar esta última fase:

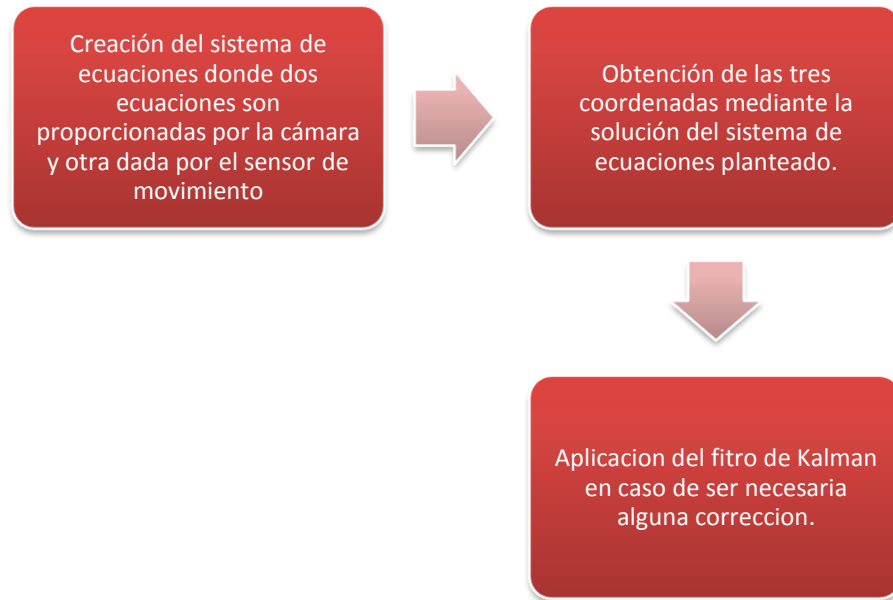


Figura 51. Diagrama de flujo del cálculo de la posición 3D

De la fase de calibración (para formar la matriz de proyección) se obtiene que cada correspondencia entre un punto en el espacio 3D y el punto en el espacio 2D de la pantalla de la cámara genera dos ecuaciones lineales basadas en los elementos de la matriz de proyección.

$$x_i = \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}} \quad (23)$$

$$y_i = \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}} \quad (24)$$

Obteniendo el centro de la persona objetivo como punto de referencia para la obtención de la posición real y sustituyendo los valores numéricos de los elementos de la matriz de proyección que se obtuvieron en la primera fase del proyecto, simplemente sería necesario obtener la ecuación extra que establecerá el sistema de ecuaciones.

De la distancia de un punto a otro en el espacio 3D, la última ecuación del sistema es proporcionada por el sensor de distancia y se representa de la siguiente forma:

$$d = \sqrt{(X_{persona} - X_{camara})^2 + (Y_{persona} - Y_{camara})^2 + (Z_{persona} - Z_{camara})^2} \quad (30)$$

Los valores de X_{camara} , Y_{camara} y Z_{camara} son conocidos y representan la posición en el espacio 3D que se encuentra la cámara.

Para resolver el sistema de ecuaciones de tres variables con tres incógnitas, se usará el comando "solve" de MatLAB que resuelve un sistema de ecuaciones previamente dadas por parámetro.

En las siguientes ilustraciones se puede observar los resultados obtenidos simplemente con la detección.



Figura 52. Detección realizada

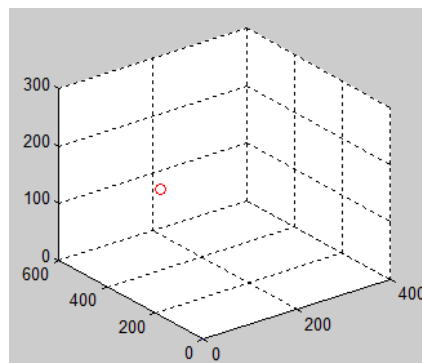


Figura 53. Punto en 3D correspondiente a la detección de la figura 40

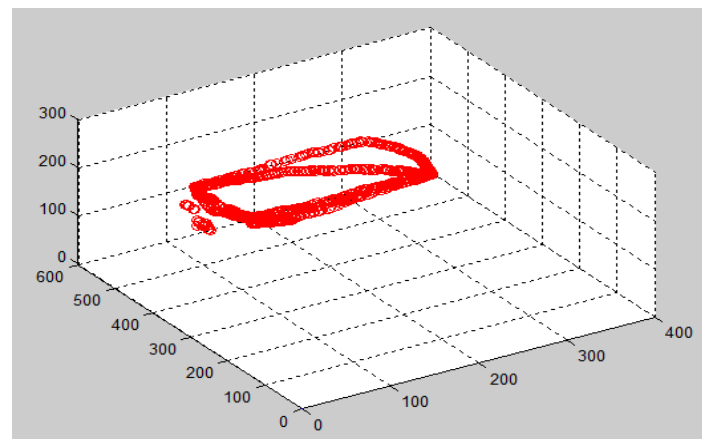


Figura 54. Trayectoria en 3D obtenida

En este caso se puede comprobar cómo hay algunos cálculos de las coordenadas que se salen de lo lógico.

Para poder solventar este problema, se aplicará un filtrado de Kalman.

Filtro de Kalman

El filtro de Kalman es un algoritmo que se basa en el modelo de espacio de estados de un sistema para estimar el estado futuro y la salida futura realizando un filtrado óptimo a la señal de salida, y dependiendo del retraso de las muestras que se le ingresan puede cumplir la función de estimador de parámetros o únicamente de filtro. Pero en ambos casos elimina ruido, estas ecuaciones son ampliamente utilizadas ya que incluyen probabilidades estadísticas puesto que toma en cuenta la aleatoriedad tanto de la señal como del ruido.

Sus ecuaciones solo dependen de una muestra anterior y la muestra presente lo que permite un ahorro considerable de memoria a la hora de ser implementado en un sistema digital y su fácil programación lo hace muy atractivo ya que se basa en un método recursivo.

Entre varias de sus aplicaciones se encuentran la estimación demográfica, procesamiento de señales biológicas, sistemas de navegación, predecir el comportamiento de variables económicas, procesamiento de imágenes, entre otras.

Debido a su gran campo de acción se hace muy importante conocer su funcionamiento para así tener las herramientas básicas que permitan la solución de diversos problemas prácticos de forma sencilla y óptima.

Con el filtro de Kalman, el objetivo será la obtención de un estimador óptimo de un sistema dinámico (movimiento de la persona) basado en observaciones ruidosas (cálculo de la posición donde se está usando sistemas no perfectos como la matriz pseudo inversa y fallos en la detección) y en un modelo de la incertidumbre de la dinámica del sistema.

Entre varias aplicaciones se encuentra la estimación demográfica, procesamiento de señales biológicas, sistemas de navegación, predecir el comportamiento de variables económicas, procesamiento de imágenes, estimación de canales de comunicaciones entre otras.

La solución es óptima mientras el filtro combine toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado de tal manera que el error es minimizado estadísticamente. [31]

En el caso que se estudia en este proyecto, se podría estimar que las personas se mueven a una velocidad máxima de cuatro kilómetros por hora, que ayudará en la

estimación de la siguiente posición en la que se encuentra el objetivo y poder corregir posibles falsos positivos.

Algoritmo del filtro de Kalman discreto

El filtro de Kalman estima un proceso usando una forma de realimentación y control: el filtro estima el estado del proceso en un tiempo y después obtiene la realimentación en forma de mediciones (ruidosas).

Las ecuaciones del filtro de Kalman caen en dos grupos: ecuaciones de actualización de tiempo y ecuaciones de actualización de mediciones. Las ecuaciones de actualización de tiempo son responsables de proyectar hacia adelante (en el tiempo) el estado actual y estimaciones de error y covarianza para obtener los estimados a priori del siguiente paso en el tiempo. Las ecuaciones de actualización de medida pueden ser vistas como ecuaciones correctoras.

De hecho el algoritmo final de estimación reúne los algoritmos de predicción-corrección para la resolución de problemas numéricos.



Figura 55. Funcionamiento del filtro de Kalman

Las ecuaciones específicas para actualización de tiempo y medida son presentadas en las tablas 4 y 5.

$$\hat{X}_{\bar{k}} = A\hat{X}_{k-1} + Bu_k \quad (31)$$

$$P_{\bar{k}} = AP_{k-1}A^T + Q \quad (32)$$

Tabla 4. Ecuaciones de actualización de tiempo del filtro de Kalman discreto.

De nuevo nótese como la actualización de tiempo en la tabla 4 proyecta el estado y la covarianza estimados adelante en el tiempo del paso $k - 1$ al paso k . A y B son matrices del modelo de estados mientras que Q es la varianza del ruido del proceso.

$$K_k = P_{\bar{k}}H^T(HP_{\bar{k}}H^T + R)^{-1} \quad (33)$$

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k(z_k - H\hat{x}_{\bar{k}}) \quad (34)$$

$$P_k = (1 - K_kH)P_{\bar{k}} \quad (35)$$

Tabla 5. Ecuaciones de actualización de mediciones del filtro de Kalman discreto

Por último para la siguiente iteración se debe hacer una reasignación de las variables, como se muestra en la tabla 6.

$\hat{x}_{\bar{k}} = \hat{x}_k$ $P_{\bar{k}} = P_k$

Tabla 6. Ecuaciones de reasignación de variables

La primera tarea durante la actualización de medidas es el cálculo de la ganancia de Kalman, J . El siguiente paso es medir el proceso para obtener B , y después generar un estimado del estado a posteriori incorporado a la medición. El paso final es la obtención de una covarianza de error estimada a posteriori.

Después de cada par de actualizaciones de tiempo y mediciones, el proceso es repetido con la estimación previa a posteriori usado para proyectar o predecir la nueva estimación a priori. Esta naturaleza recursiva es una de las buenas cualidades del filtro de Kalman, lo que hace prácticas muchas implementaciones.

El filtro de Kalman a pesar de sus condiciones recursivas basa su estimado actual en todas las mediciones pasadas. La figura 55 ofrece un esquema completo de la operación del filtro [32].

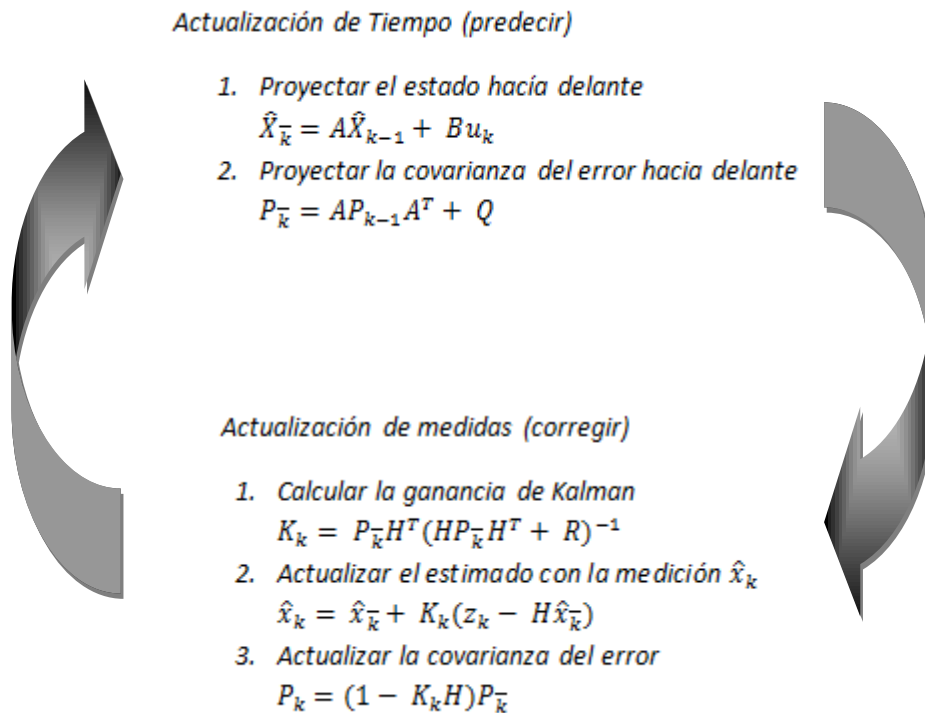


Figura 56. Esquema completo de la operación del Filtro de Kalman.

Siendo P_k^- el estimado de la covarianza del error a priori y R_k la covarianza del error medio.

En las siguientes ilustraciones se puede comprobar cómo el filtro de Kalman estima la trayectoria y nos permite corregir los errores de detección del paso anterior.

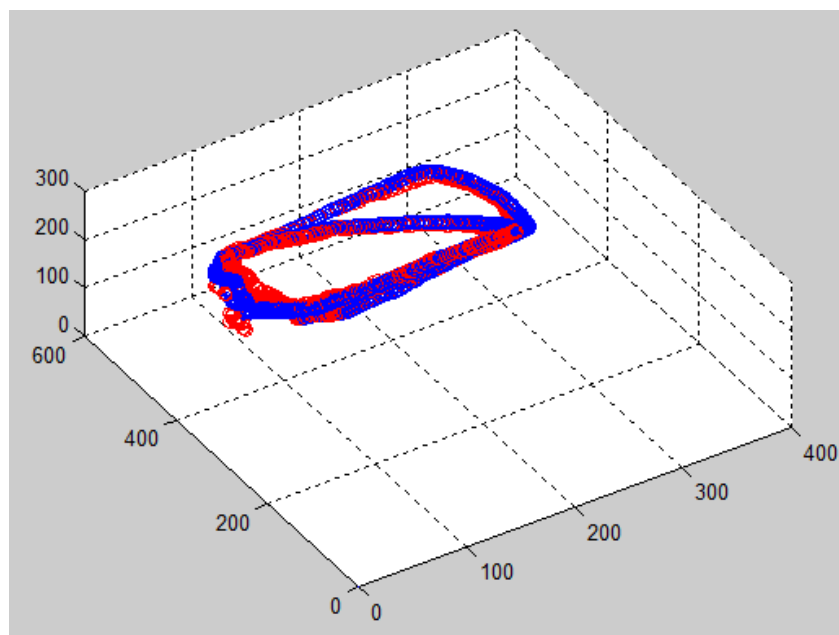


Figura 57. Uso del filtro de Kalman

La trayectoria original (color rojo) es corregida en los puntos en los que diverge y como se puede comprobar en la trayectoria estimada por el filtro de Kalman (azul), se obtiene una trayectoria más uniforme y continua.

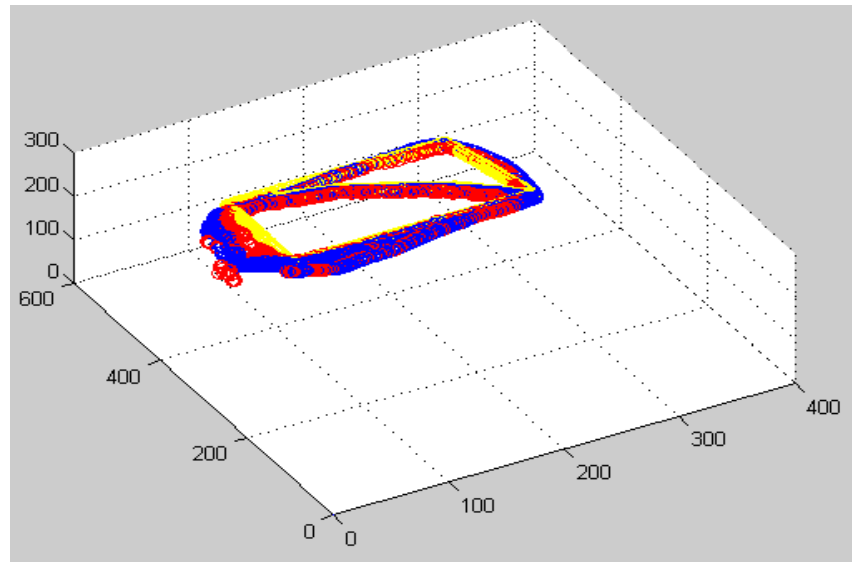


Figura 58. Representación de la trayectoria ideal, trayectoria obtenida con y sin filtro de Kalman

- La trayectoria roja representa los valores calculados directamente de la detección de la persona.
- La trayectoria azul representa los valores de trayectoria calculados del paso anterior y añadiéndole el filtro de Kalman
- La trayectoria amarilla son los valores exactos de la trayectoria seguida por la persona del video usado.

Capítulo 4: ESTUDIO Y EVALUACIÓN DE LOS RESULTADOS

4.1 Calibración

Para calcular el rendimiento de este algoritmo y uso de la matriz pseudo inversa, se realiza un test en el que se comparan los resultados obtenidos al seleccionar a mano el pixel que correspondía a uno de los puntos de calibración 3D y al resultado que nos daba haciendo uso de la matriz de proyección calculada y ese mismo punto 3D usado para su cálculo.

Se ha obtenido en todos los test usados un error del 0.1% a la hora de calcular el pixel haciendo uso de la matriz de proyección directamente.

Este resultado es muy satisfactorio teniendo en cuenta que además, es posible realizar un post procesado sobre la matriz de rotación y vector de traslación que podría mejorar el rendimiento y precisión de la matriz de proyección

Puesto que el margen de mejora es pequeño y el post procesado implicaría aumento de tiempo y coste de procesado, se decide no llevar a cabo esa pequeña mejora.

4.2 Segmentación del fondo

4.2.1 Mixture of Gaussians background extraction

Este algoritmo tenía un gran problema y es que es necesario hacer un procesado pixel a pixel que ralentiza mucho su tiempo de ejecución y procesamiento.

El resultado final es bastante bueno aunque no suficiente por si solo como se puede apreciar en la figura de a continuación.

Este algoritmo además planteó un grave problema de “ocultamiento” cuando la persona llevaba vestimenta que concordara

con el color del fondo y que por tanto el algoritmo pensase que pertenecía a este.

También, como se puede apreciar, la imagen fue reducida de 240x320 a 60x80 ya que si no, el algoritmo se hacía muy lento de probar.



Figura 59. Rendimiento del algoritmo

En general, las pruebas que se realizaron dieron como resultado que no era suficiente este algoritmo pos si solo para conseguir mantener una detección suficientemente buena de la figura en casi toda la trayectoria que nos permitiera clasificar esa figura como persona o no.

4.2.2 Edge background extraction

Para realizar la clasificación, se decidirá si un pixel pertenece a un objeto en movimiento si se tiene un nivel de confianza del 90% o mayor. Este valor mínimo de confianza ha de ser escogido de forma que se obtenga un rendimiento deseado mínimo y por tanto es personalizable. A su vez, las constantes M y m fueron escogidas con los valores 12 y 6 respectivamente.



Figura 60. Rendimiento de este algoritmo

En cambio, juntando ambos algoritmos, vemos como mejora la situación.

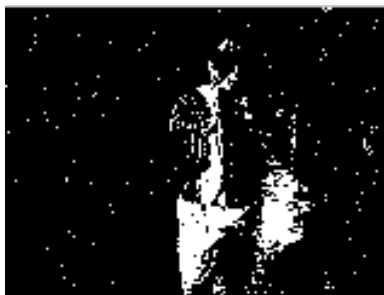


Figura 61. Rendimiento de algoritmo basado en Gaussian

4.2.3 Luminance and chrominance distortion background extraction

Este algoritmo es el que mejor resultado ha dado. Tiene la gran ventaja que es posible realizar mucho procesamiento por matrices lo que mejora mucho la velocidad de procesamiento y se ha sido capaz de obtener un gran resultado usando solo este algoritmo.

Cabe destacar que para probar la robustez de este algoritmo se fue modificando la intensidad de luz durante la fase de entrenamiento y además, en la fase de detección, incluimos un haz de luz muy intenso a ver como se comportaba y aun así, se obtuvieron unos resultados muy buenos.

Se consiguieron unos resultados muy satisfactorios a pesar de que en algunos momentos se tuviera alguna detección que no fuera la persona en movimiento pero que podría ser perfectamente eliminada en la fase de clasificación de personas ya que no se identificaría esa detección como persona.

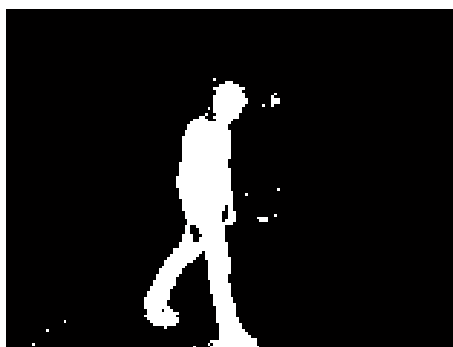


Figura 62. Rendimiento del algoritmo basado en luminancia y croma

❖ Toma de decisión

Finalmente, para la fase de detección del fondo es necesario decidir entre el algoritmo de mezcla de gaussianas y el algoritmo basado en la distorsión de la luminancia y de la croma ya que el algoritmo de detección de bordes ayuda a la detección de una silueta más definida pero por si solo no da rendimiento suficiente y podría usarse como ayuda en cualquier caso.

Como se muestra a continuación, para tres frames determinados se han obtenido el % de pixeles erróneos y tiempo de entrenamiento necesario para cada algoritmo. Además, se incluye el rendimiento del algoritmo solo de detección de variación de la magnitud de bordes para concluir que por sí solo no sería suficiente para obtener una detección de personas lo mínimamente necesario.

Los tres frames que se han usado son los siguientes:



Figura 63. Imágenes referencia cálculo rendimiento

Y las tres plantillas de detección perfecta que se han realizado son las siguientes:

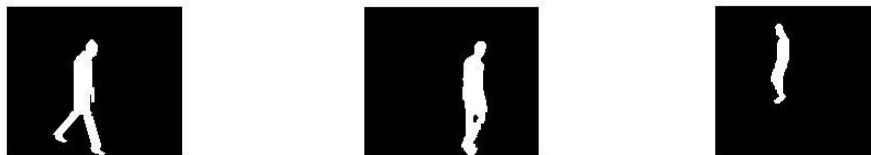


Figura 64. Plantillas referencia cálculo de rendimiento

	Píxeles erróneos frame 107	Píxeles erróneos frame 150	Píxeles erróneos frame 271	Media	Dificultad de entrenamiento
Edge background extraction sin post procesado	6,78%	7,02%	3,2%	5,66%	1
Edge background extraction con post procesado	6,58%	9,1%	3,48%	6,38%	1
Luminance and chrominance distortion background extraction + edge background extraction sin post procesado	3,08%	1,51%	1,77%	2,12%	4
Luminance and chrominance distortion background extraction + edge background extraction con post procesado	2,79%	1,56%	1,51%	1,95%	4
Mixture of Gaussians background extraction + edge background extraction sin post procesado	8,30%	7,97%	4,32%	6,86%	20
Mixture of Gaussians background extraction + edge background extraction con post procesado	6,39%	6,91%	2,89%	5,39%	20

Tabla 7. Rendimiento de los algoritmos de segmentación del fondo

El parámetro Dificultad representa cómo de pesado procesalmente es el algoritmo. Cuanto mayor es la Dificultad, más tiempo tarda en entrenarse el algoritmo

y más consumo de procesamiento tiene la evaluación de cada frame. Cada nivel de dificultad representa en este caso dos minutos.

Teniendo en cuenta la tabla anterior, se puede concluir que el algoritmo basado en la distorsión de la luminancia y la croma junto con el algoritmo de detección de variación de la magnitud y ángulo de bordes y post procesado da el mejor resultado tanto en el error de píxeles en la mayoría de los frames como por su tiempo de ejecución bastante bajo.

Como se puede comprobar en las siguientes imágenes, se ha obtenido como resultado la detección de objetos con una silueta muy definida que permitirá en la siguiente fase realizar una clasificación muy precisa y así descartar objetos en movimiento que no nos son de interés, es decir, no son personas.



Figura 65. Rendimiento total fase detección fondo 1



Figura 66. Rendimiento total fase detección fondo 2

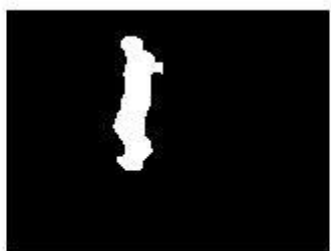


Figura 67. Rendimiento total fase detección fondo 3



Figura 68. Rendimiento total fase detección fondo 4

4.3 Clasificación de personas

En este apartado se ha podido comprobar cómo se es capaz de distinguir entre los distintos objetos detectados y saber cuáles son personas o no de forma bastante precisa.

Se ha sido capaz además de saber el número de personas que se encuentran en la habitación que puede servir a la hora de incluir mejoras.

En las siguientes tres imágenes se puede comprobar que la clasificación se hace de forma perfecta incluso cuando sombras son detectadas como objetos en movimiento además de un indicador con el número de personas en ese frame.

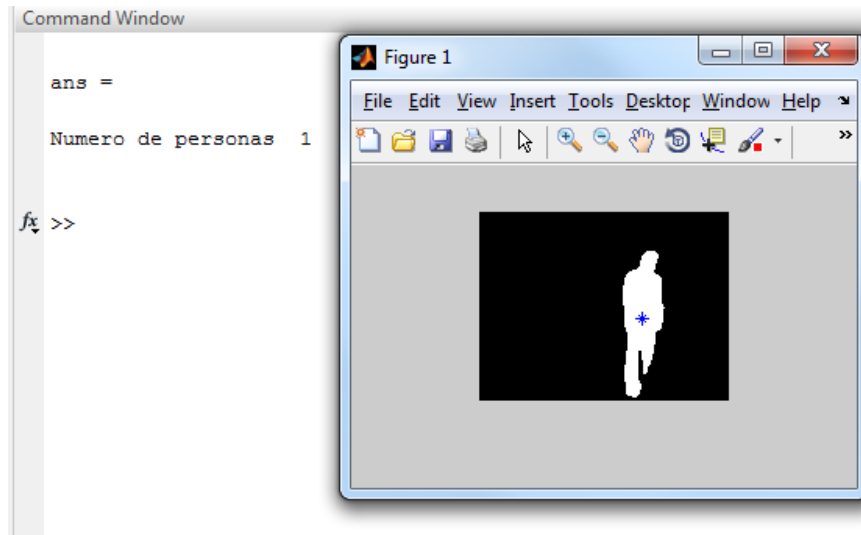


Figura 69. Clasificación de una persona sin otros elementos detectados

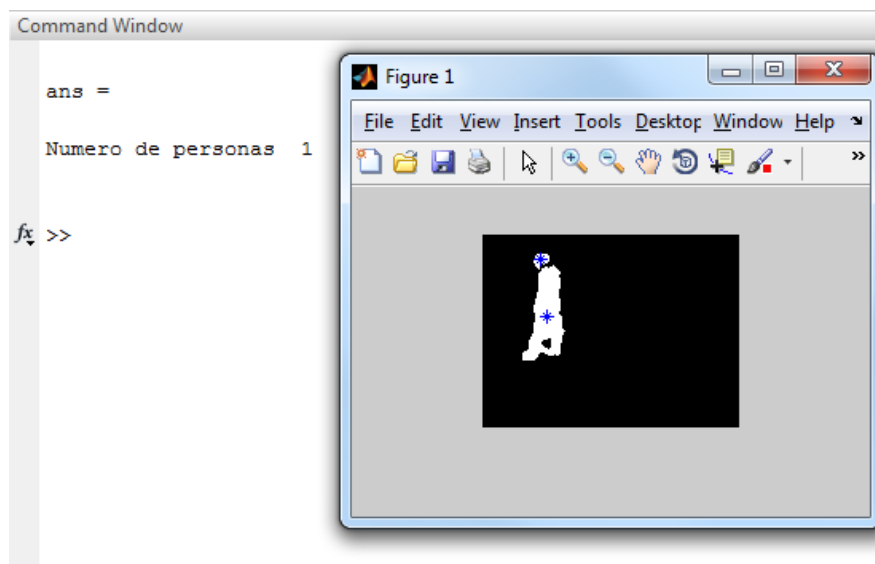


Figura 70. Clasificación de una personas con más elementos detectados

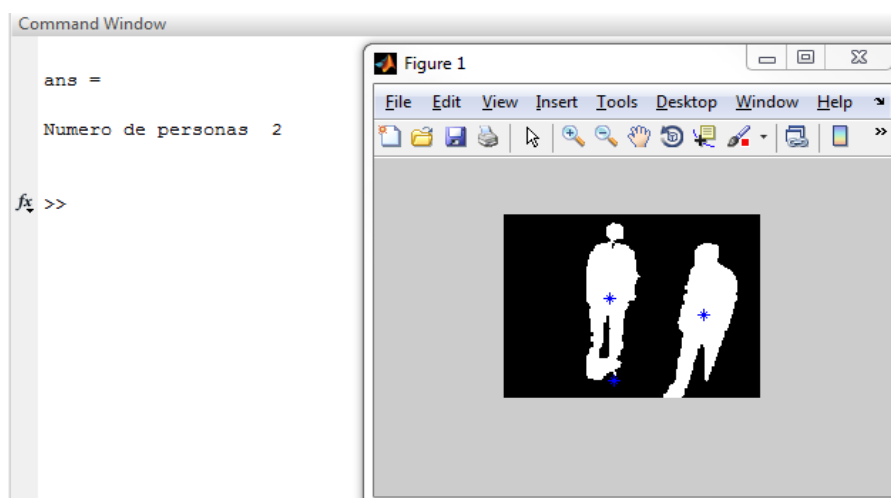


Figura 71. Clasificación de dos personas con otros elementos detectados

4.4 Fusión de sensores

En esta sección, vamos a hacer un estudio del rendimiento de la aplicación completa. Se comprobará también cómo el algoritmo de Kalman ayuda a corregir las clasificaciones erróneas obtenidas. Analizando el elemento extra, el sensor de movimiento, se va a exponer un estudio de cómo afecta la recepción incorrecta de la distancia por parte del sensor al cálculo de la trayectoria.

A continuación se exponen unos resultados que muestran el error en mínimos cuadrados a lo largo de toda la trayectoria.

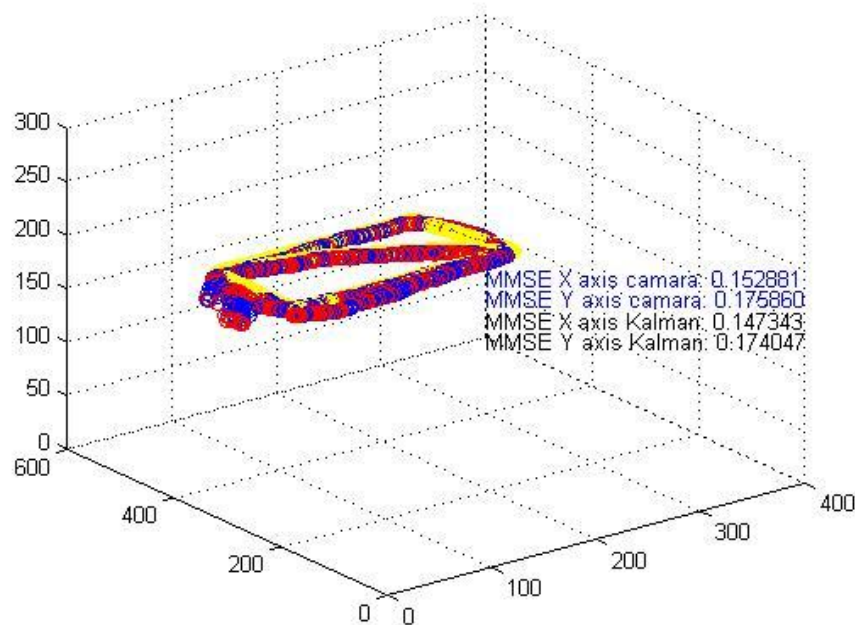


Figura 72. Trayectoria obtenida y error en mínimos cuadrados

Los dos primeros valores que se muestran en azul representan el error (en mínimos cuadrados) de la detección de la cámara (trayectoria roja) con respecto a la trayectoria ideal (trayectoria amarilla) para la coordenada X y la coordenada Y respectivamente.

Los dos valores que se muestran en negro representan el error (en mínimos cuadrados) de la trayectoria estimada con el filtro de Kalman (trayectoria azul) con respecto a la trayectoria ideal (trayectoria amarilla) para la coordenada X y la coordenada Y respectivamente.

Todos los valores mostrados se miden en metros y teniendo en cuenta que se han medido variaciones máximas de 1,5 metros en el eje X y 2,1 metros en el eje Y a lo

largo de toda la trayectoria, se ha tenido un error de 1,5% en el eje X y 1,47% en el eje Y .

El error obtenido tanto para el eje X como para el eje Y en el caso del filtro de Kalman es más pequeño ya que, en el caso de cálculos de posiciones erróneas, el filtro de Kalman estima una posición (basándose en los parámetros que hemos definido en él previamente), una medición mucho más cercana a la real.

Para analizar el caso de que el sensor de distancia no reporte correctamente la distancia a la que está el objetivo, se ha llevado a cabo una simulación en la que se va sumando valores aleatorios a la medida de la distancia real y ver cómo afecta al error total sobre la trayectoria. Se ha introducido un desfase de entre 0 y 50 centímetros distribuidos en la siguiente forma: entre 0 y 5 centímetros, entre 0 y 10 centímetros, entre 0 y 15 centímetros, etc.

Como se puede comprobar en la siguiente imagen, el error medio calculado entre la trayectoria ideal y las trayectorias calculadas con y sin filtro de Kalman es menor haciendo uso del filtro de Kalman que sólo haciendo uso del cálculo de cada posición (resolviendo el sistema de tres ecuaciones con tres incógnitas planteado). Además, los cálculos erróneos de la posición que provocarían comportamientos que no tendrían sentido (como desplazarse entre un frame y otro seis metros por ejemplo) no son tenidos en cuenta si se hace el filtro de Kalman (ya que estimaría una posición mucho más cercana a la real). Basándose en los resultados obtenidos en estas pruebas, se puede concluir que el uso del filtro de Kalman mejora el rendimiento total del sistema y nos permite crear una trayectoria más real y más cercana a la real.

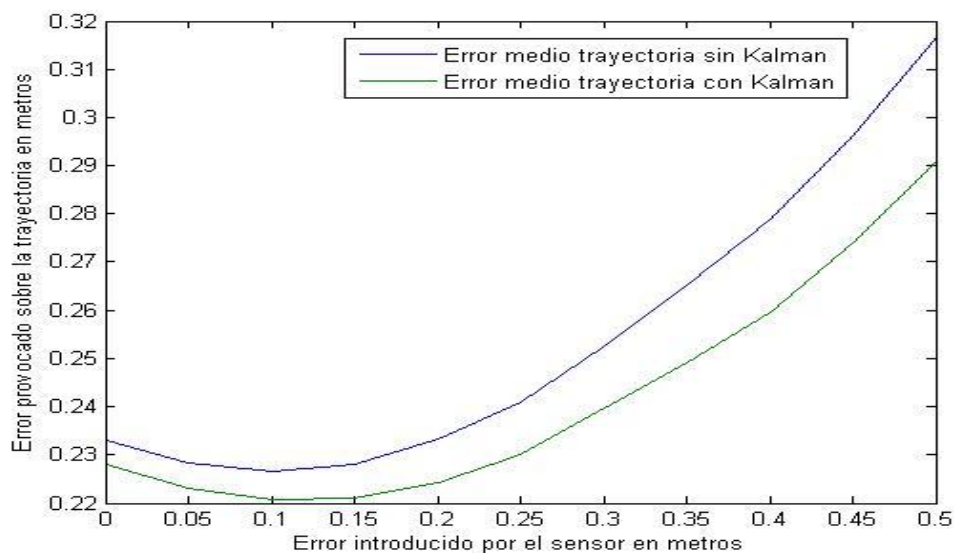


Figura 73. Estudio de las consecuencias del error introducido por el sensor de distancia

Puesto que el objetivo de este proyecto es obtener la trayectoria para poder obtener pautas de comportamiento, el pequeño error obtenido no modifica el objetivo para lo que se ha hecho un estudio en este proyecto y sigue permitiendo obtener todas las características que se quieren de la trayectoria calculada.

Capítulo 5: CONCLUSIÓN

A lo largo de este trabajo se presenta una solución para el cálculo de trayectorias de personas y que sirva de punto de partida para múltiples aplicaciones. Hace uso de una estructura establecida muy básica, una cámara y un sensor de distancia, que junto al uso de algunos algoritmos, reportan un resultado muy satisfactorio.

Para desarrollar esta tarea se han seguido una serie de fases en cascada: calibración, segmentación del fondo, clasificación de personas y extracción de la coordenada 3D. Aquella que ha requerido mayor esfuerzo ha sido claramente la segmentación del fondo.

Se ha puesto especial énfasis en la fase de segmentación del fondo puesto que, aunque todas las fases son importantes y sin ninguna de ellas se podría haber llevado a cabo esta solución, era especialmente crítica al necesitar una detección de la silueta de una persona lo más definida posible para así, poder clasificar después en la siguiente fase, el objetivo detectado como una persona (por tanto de interés) o cualquier otra cosa (de no interés).

El primer objetivo fue conseguir calibrar la cámara y así ser consciente de la perspectiva de las distancias del lugar de aplicación.

Calibrar la cámara consiste en obtener una matriz que relaciona un punto en el espacio 3D (espacio real) con un punto en el espacio 2D (pantalla de la cámara) que se denomina matriz de proyección.

Para realizar la calibración, es necesario obtener un número mínimo de relaciones de los espacios 3D->2D conocidas, es decir, puntos en los que se sepan tanto sus coordenadas 3D como sus coordenadas 2D. El número mínimo de puntos conocidos para obtener una matriz de proyección suficientemente buena son seis [24]. Esto es debido a que con seis puntos se reduce hasta valores muy pequeños la posibilidad de que haya combinaciones lineales entre esos puntos y, por tanto, una matriz de proyección errónea.

La matriz de proyección se obtiene de resolver la ecuación **(3)**. La matriz X de esta ecuación está formada por los valores de las coordenadas de los puntos en el espacio 3D y sus relaciones con el espacio 2D conocidos que se ha indicado anteriormente.

Para poder despejar el vector cuyos valores ordenados forman la matriz de proyección que se quiere obtener es necesario, en la ecuación **(3)**, despejar el vector C . Por lo tanto, se necesita pasar al otro lado de la igualdad la matriz X . Como esta matriz no es cuadrada (mismo número de columnas y de filas), no se puede realizar la inversa. Para ello, se hace uso de la matriz pseudo inversa (ver **Anexo**) para resolver el sistema.

Para mejorar el rendimiento de esta fase, se propone un procesamiento posterior sobre la matriz de proyección. Para ello se separa esta matriz de proyección en un vector de traslación y una matriz de rotación y se trabaja sobre estos dos elementos por separado para aumentar la precisión de la matriz después de formarse de nuevo.

Para la segmentación del fondo, se han llevado a cabo tres algoritmos diferentes, Mixture of Gaussians background extraction, Illuminance and Chrominance background extraction y Edge background extraction.

El algoritmo Mixture of Gaussians para la segmentación del fondo consiste en la caracterización de cada pixel mediante entre tres o cinco gaussianas (dependiendo de la capacidad de procesamiento que se tenga) que contengan los valores típicos de color que toma cada pixel y así detectar cuando un pixel tiene un valor diferente, y por tanto, que no pertenezca al fondo.

Durante el entrenamiento de este algoritmo, se comprueba si cada una de las gaussianas usadas coincide, para cada pixel de la imagen, con el valor actual de este. Si alguna de esas gaussianas ha detectado ese pixel como del fondo, aumenta su importancia a la hora de decidir si ese pixel pertenece al fondo o no y actualiza su valor teniendo en cuenta el valor actual del pixel y el que ya tenía la gaussiana. Además, el resto de gaussianas reducirán su peso sobre ese pixel.

Después del entrenamiento, sobre cada pixel tendremos una o más gaussianas que lo identifican como del fondo y por tanto, detectar movimiento.

El algoritmo de Illuminance and Chrominance para la segmentación del fondo hace uso de los componentes de la luz, la intensidad del brillo y la croma, para comprobar cuando un objeto pertenece al fondo o no.

Una vez que se ha obtenido una imagen de referencia en la fase de entrenamiento donde se obtiene unos valores de intensidad de brillo y croma para cada pixel, se pasa a la fase de aplicación, donde se calcula la distorsión del brillo y la distorsión de la croma entre la imagen de referencia y la imagen en evaluación y teniendo en cuenta unos umbrales definidos, se es capaz de saber diferentes tipos de objetos como fondo destacado, objeto en movimiento, sombra, etc.

Por último, el tercer algoritmo para la segmentación del fondo que se ha evaluado ha sido el Edge background extraction. En este caso, se hace un estudio de los bordes de los objetos que forman el fondo (tal como mesas, sillas, vértices de paredes, etc.) y se obtienen dos imágenes de referencia, una para los bordes verticales y otra para los bordes horizontales.

En la fase de detección se obtienen los bordes (tanto horizontales como verticales) de la imagen actual y se calcula en cuanto han variado en magnitud y en ángulo esos bordes con respecto a los de referencia, que se han obtenido previamente en la fase de entrenamiento. Por último y basándose en el cálculo anterior, se obtiene con qué probabilidad no pertenecen al fondo esos bordes detectados, y por tanto, en movimiento.

Después de aplicar esos algoritmos, en todos los casos se ha realizado un procesamiento posterior para obtener siluetas lo mayor definida posibles y que ayudara en la siguiente fase de clasificación de los objetos detectados. Para ello, se hace uso de la erosión y dilatación como procesos para eliminar zonas aisladas de píxeles y rellenar zonas con pequeños huecos.

Para saber cuáles son los mejores algoritmos y hacer uso de ellos, se ha llevado a cabo un estudio del porcentaje de píxeles erróneos que ha tenido cada algoritmo. Se crean tres plantillas para tres frames diferentes de una trayectoria que corresponderían a una detección perfecta y se calcula la diferencia entre la plantilla y la detección hecha por el algoritmo.

Esta evaluación llevada a cabo en el presente proyecto se puede encontrar en las tabla 7.

Para la tercera fase, la clasificación de los objetivos detectados, se ha llevado a cabo un estudio morfológico de la persona en la que se observa que tiene unas características, por la forma de su cuerpo, que lo distinguen de casi cualquier cosa.

Este estudio consiste en obtener dos proyecciones, una horizontal y otra vertical, de los puntos que forman el objeto detectado y compararlas con una proyección vertical y otra horizontal que previamente se han construido a partir de detecciones conocidas de personas. Estas proyecciones identifican de forma muy particular a los humanos por su estructura corporal.

Las figuras 47 y 48 representan cómo son estas proyecciones.

En esta parte de clasificación se propone una mejora que permitiría saber en qué posición (tumbado, sentado, inclinado, mirando hacia delante, etc.) se encuentra esa persona y por tanto obtener mucho mas información sobre su comportamiento.

Esta identificación extra de la postura es posible obtenerse siguiendo la misma filosofía seguida anteriormente ya que, en función de ligeras variaciones de alguna de las dos proyecciones, se puede identificar en qué postura se encuentra el objetivo en ese momento.

También es posible identificar ciertas partes del cuerpo segmentándolo mediante figuras volumétricas (reconstrucciones de partes del cuerpo con las que se puedan comparar las detectadas) que permitan, en función de donde se encuentran esas partes del cuerpo identificadas, estimar la postura que tiene el objetivo.

La última fase que se ha llevado a cabo ha sido la fusión de todos los sensores y así obtener la posición 3D del objetivo basándonos como punto de referencia 2D de la cámara, el centro de la figura humana detectada. En este momento del proyecto es cuando entra en acción el sensor de distancia que lleva consigo el objetivo.

Se plantea un sistema de tres ecuaciones, dos proporcionadas por la matriz de proyección y otra extra de la ecuación de la distancia entre un punto y otro en el espacio 3D, y tres incógnitas, las tres coordenadas en el mundo 3D real.

Resolviendo este sistema de ecuaciones, se obtiene las coordenadas 3D del objetivo. Además, se realiza un filtrado de Kalman, que estima la posición futura del objetivo basándose en la posición anterior y parámetros establecidos que identifican el movimiento de una persona como la velocidad, aceleración, etc. El filtro de Kalman nos permite corregir posibles cálculos de coordenadas erróneos y que no tengan sentido a lo largo de una trayectoria además de dar una trayectoria más cercana a la ideal.

REFERENCIAS

- [1] **I. Haritaoglu, D. Harwood y L. S. Davis**, «Real-Time Surveillance of People and Their Activities,» de *IEEE Transactions on pattern analysis and machine intelligence*, 2000.
- [2] **S. Sternig, T. Mauthner, A. Irschara, P. M. Roth y H. Bischof**, «Multi-camera Multi-object Tracking by Robust Hough-based,» Institute for Computer Graphics and Vision, Graz University of Technology.
- [3] **J. Cui, H. Zha, H. Zhao y R. Shibasaki**, «Multi-modal tracking of people using laser scanners and video camera,» 2007.
- [4] **Z.Zhang**, «Camera Calibration,» de *Emerging Topics in Computer Vision*, 2004.
- [5] **D. C. Brown**, «Close-range camera calibration».
- [6] **W. Faig**, «Calibration of close-range photogrammetry systems: Mathematical formulation,» de *Photogrammetric Engineering and Remote Sensing*.
- [7] **S. Ganapathy**, «Decomposition of transformation matrices for robot vision.,» de *Pattern Recognition Letters*.
- [8] **D. Gennery**, «Stereo-camera calibration,» de *Proceedings of the 10th Image Understanding Workshop*.
- [9] **R. Y. Tsai**, «A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses,» de *IEEE Journal of Robotics and Automation*.
- [10] **O. Faugeras**, «Three-Dimensional Computer Vision: a Geometric Viewpoint,» de *MIT Press*.
- [11] **Z.Zhang**, «A flexible new technique for camera calibration,» de *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [12] **P. F. Sturm y S. J. Maybank**, «On plane-based camera calibration: A general algorithm, singularities, applications.,» de *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [13] **Z. Zhang**, «Camera calibration with one-dimensional objects.,» de *Proc. European Conference on Computer Vision* .

- [14] **S. Brutzer, B. Hoferlin y G. Heidemann**, «Evaluation of Background Subtraction Techniques for Video Surveillance,» Intelligent Systems Group, Universitat Stuttgart, Germany.
- [15] **Y. Dedeoglu**, «Moving Object Detection, Tracking and Classification for Smart Video Surveillance,» 2004.
- [16] **A. Lipton, H. Fujiyoshi y R.S.Patil**, «Moving target classification and tracking from real-time video,» de *Proc. of Workshop Applications of Computer Vision*, 1998, pp. 129-136.
- [17] **R.T.Collins**, «A system for video surveillance and monitoring: VSAM final report».
- [18] **T. Brodsky**, «Visual Surveillance in Retail Stores and in the Home,» de *Video-Based Surveillance Systems*, Boston, 2002, pp. capitulo 4, 51-61.
- [19] **C. Papageorgiou, T. Evgeniou y T.Poggio**, «A trainable pedestrian detection system,» de *Proc. of IEEE Int. Conf. on Intelligent Vehicles*, Germany, 1998, pp. 241-246.
- [20] **S. Johnsen y A. Tews**, «Real-Time Object Tracking and Classification Using a Static Camera,» Japon, 2009.
- [21] **R. Cutler y L.S.Davis**, «Robust real-time periodic motion detection, analysis and applications,» de *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, pp. 781-796.
- [22] **Y. Bar-Shalom y W. D. Blair**, «Multitarget multisensor tracking: applications and advances Volume III,» 200.
- [23] **L. Snidaro, G. L. Foresti, R. Niu y P. K. Varshney**, «Sensor fusion for video surveillance,» de *Electrical Engineering and Computer Science*, 2004.
- [24] **C. R. Viala, A. J. S. Salmerón y R. S. Fernández**, «ERRORES EN LA ESTIMACIÓN DE LA MATRIZ DE PROYECCIÓN,» Departamento de Ingeniería de Sistemas y Automática , Universidad Politécnica de Valencia.
- [25] **G.-Q. Wei y S. D. Ma**, «Implicit and Explicit Camera Calibration:Theorv and Experiments,» de *EEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 1994.
- [26] **W. G. Chris Stauffer**, «Adaptive background mixture models for real-time tracking,» The Artificial Intelligence Laboratory, Massachusetts Institute of Technology.
- [27] **T. Horprasert, D. Harwood y L. S. Davis**, «A Statistical Approach for Real-time Robust,» de *Computer Vision Laboratory*, University of Maryland.
- [28] **O. Javed, K. Shafique y M. Shah**, «A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information,» 2002.

- [29] **U. d. Sevilla**, «Departamento de Matemática aplicada, Tema 3: Filtros,» 2013. [En línea]. Available: <http://grupo.us.es/gtocom/pid>. [Último acceso: 2014].
- [30] **U. d. Sevilla**, «Departamento de Matemática Aplicada I, Tema 5: Morfología,» 2013. [En línea]. Available: <http://grupo.us.es/gtocom/pid>. [Último acceso: 2014].
- [31] **Á. S. Ramírez**, *EL FILTRO DE KALMAN*, DEPARTAMENTO DE INVESTIGACIONES ECONÓMICAS.BANCO CENTRAL DE COSTA RICA, 2003.
- [32] **J. A. C. Cárdenas, M. A. N. Arias y V. A. O. Bravo**, ANÁLISIS Y APLICACIÓN DEL FILTRO DE KALMAN A UNA SEÑAL CON RUIDO ALEATORIO, Ingeniería electrónica, Universidad Tecnológica de Pereira, Pereira, Colombia, 2013.
- [33] **A. Yildiz y Y. S. Akgul**, A Fast Method for Tracking People with Multiple Cameras, Vision Lab. Gebze Institute of Technology.

ANEXO

Matriz pseudoinversa

La pseudo-inversa de una matriz $m \times n$, A es una matriz que generaliza a matrices arbitrarias la noción de la inversa de una matriz cuadrada, matriz invertible. El pseudo-inversa puede expresarse a partir de la descomposición de valor singular (SVD) de A , como sigue.

Se definen los valores singulares de A como:

$$A = U \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} V^T$$

donde U , V son ambas matrices ortogonales, y S es una matriz diagonal que contiene los (positivos) valores singulares de A en su diagonal.

A continuación, la pseudo-inversa de A es la matriz $n \times m$ definida como

$$A^T = U \begin{pmatrix} S^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T$$

Tenga en cuenta que A^T tiene la misma dimensión que la traspuesta de A .

Esta matriz tiene muchas propiedades útiles:

- Si A es de rango de columna completo, significando $rank(A) = n \leq m$, es decir, $A^T A$ no es singular, entonces A^T es una inversa por la izquierda de A , en el sentido de que $A^T A = I_n$. Tenemos expresión de forma cerrada:

$$A^T = (A^T A)^{-1} A^T$$
- Si A es de rango de fila completa, es decir, $rank(A) = m \leq n$, es decir, AA^T es no singular, entonces A^T es una inversa por la derecha de A , en el sentido de que $AA^T = I_m$. Tenemos la expresión de forma cerrada:

$$A^T = A^T (AA^T)^{-1}$$

- Si A es cuadrada, invertible, entonces su inversa $A^T = A^{-1}$
- La solución al problema de mínimos cuadrados

$$\min_x \|Ax - y\|_2$$

y la norma mínima es $x^* = A^T y$.

